



JC408 U.S. PTO

12/03/97

## Certificate of Mailing

Date of Deposit December 3, 1997

Label Number: EI975155822US

I hereby certify under 37 CFR 1.10 that this correspondence is being deposited with the United States Postal Service as "Express Mail Post Office to Addressee" with sufficient postage on the date indicated above and is addressed to: BOX PATENT APPLICATION, Assistant Commissioner for Patents, Washington, D.C. 20231.

Sandra E. Marxen

Printed name of person mailing correspondence

Signature of person mailing correspondence

## UTILITY PATENT APPLICATION TRANSMITTAL UNDER 37 CFR §1.53(b)

Attorney Docket Number	01997/198006
Applicant	H. Robert Horvitz, Junying Yuan, Shai Shaham
Title	Cloning, Sequencing and Characterization of Two Cell Death Genes and Uses Therefor

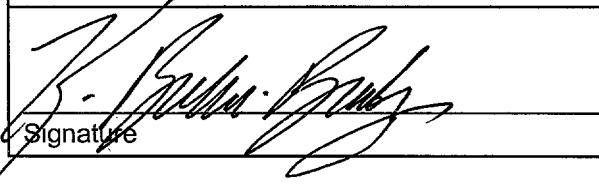
## PRIORITY INFORMATION:

This application is a continuation of and claims priority from United States patent application 08/287,669, filed August 9, 1994 which is a divisional of 07/979,638, filed November 20, 1992 which is a CIP of 07/897,788, filed June 12, 1992.

## APPLICATION ELEMENTS:

Cover sheet	01 pages
Specification	66 pages
Claims	15 pages
Abstract	01 pages
Drawing	21 pages
Combined Declaration and POA, which is: <input type="checkbox"/> Unsigned; <input checked="" type="checkbox"/> Newly signed for this application; <input checked="" type="checkbox"/> A copy from prior application 07/979,638 and the entire disclosure of the prior application is considered as being part of the disclosure of this new application and is hereby incorporated by reference therein.	03 pages
Statement Deleting Inventors	0 pages
Sequence Statement	0 pages
Sequence Listing on Paper	0 pages
Sequence Listing on Diskette	0 pages

03084-138-120233

Small Entity Statement, which is: <input type="checkbox"/> Unsigned; <input type="checkbox"/> Newly signed for this application; <input type="checkbox"/> A copy from prior application [**SERIAL NUMBER**] and such small entity status is still proper and desired.	0 pages
Preliminary Amendment	0 pages
IDS	0 pages
Form PTO 1449	0 pages
Cited References	0 pages
Recordation Form Cover Sheet and Assignment	0 pages
Assignee's Statement	0 pages
English Translation	0 pages
Certified Copy of Priority Document	0 pages
Return Receipt Postcard	1
<b>FILING FEES:</b>	
Basic Filing Fee: \$790/\$395	\$395.00
Excess Claims Fee: 68 - 20 x \$11/\$22	\$528.00
Excess Independent Claims Fee: 22 - 3 x \$82/\$41	\$779.00
Multiple Dependent Claims Fee: \$270/\$135	\$0
Total Fees:	\$1702.00
<input type="checkbox"/> Enclosed is a check for _____ to cover the total fees. <input type="checkbox"/> Charge _____ to Deposit Account No. 03-2095 to cover the total fees. <input checked="" type="checkbox"/> The filing fee is not being paid at this time. <input type="checkbox"/> Please apply any other charges, or any credits, to Deposit Account No. 03-2095.	
<b>CORRESPONDENCE ADDRESS:</b>	
Kristina Bieker-Brady, Ph.D. Reg. No. 39,109 Clark & Elbing LLP 176 Federal Street Boston, MA 02110	
Telephone: 617-428-0200 Facsimile: 617-428-7045	
 Signature	
 Date	

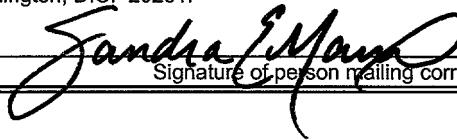
Certificate of Mailing

Date of Deposit December 3, 1997

Label Number: EI975155822US

I hereby certify under 37 CFR 1.10 that this correspondence is being deposited with the United States Postal Service as "Express Mail Post Office to Addressee" with sufficient postage on the date indicated above and is addressed to BOX PATENT APPLICATION, Assistant Commissioner of Patents, Washington, D.C. 20231.

Sandra E. Marxen  
Printed name of person mailing correspondence

  
Signature of person mailing correspondence

APPLICATION  
FOR  
UNITED STATES LETTERS PATENT

APPLICANT : H. ROBERT HORVITZ, JUNYING YUAN, SHAI SHAHAM  
TITLE : CLONING, SEQUENCING AND CHARACTERIZATION OF  
TWO CELL DEATH GENES AND USES THEREFOR

CLONING, SEQUENCING AND CHARACTERIZATION  
OF TWO CELL DEATH GENES AND USES THEREFOR

Government Funding

Work described herein was supported by grants  
5 GM24663 and GM24943 from the U.S. Public Health  
Service. The U.S. Government has certain rights in the  
invention.

Related Application

This application is a continuation-in-part of USSN  
10 07/897,788, entitled "Cloning, Sequencing and  
Characterization of Two Cell Death Genes and Uses  
Therefor" by H. Robert Horvitz, Junying Yuan, and Shai  
Shaham, filed June 12, 1992. The teachings of USSN  
07/897,788 are incorporated by reference.

15 Background

Cell death is a fundamental aspect of animal development. Many cells die during the normal development of both vertebrates (Glucksmann, *Biol. Rev. Cambridge Philos. Soc.* 26:59-86 (1951)) and invertebrates (Truman, *Ann. Rev. Neurosci.* 7:171-188 (1984)). These deaths appear to function in morphogenesis, metamorphosis and tissue homeostasis, as well as in the generation of neuronal specificity and sexual dimorphism (reviewed by Ellis et al., *Ann. Rev. Cell Biol.* 7:663-698 (1991)). An understanding of the mechanisms that cause cells to die and that specify which cells are to live and which cells are to die is essential for an understanding of animal development.

The nematode *Caenorhabditis elegans* is an appropriate organism for analyzing naturally-occurring or programmed cell death (Horvitz et al., *Neurosci. Comment.* 1:56-65 (1982)). The generation of the 959

somatic cells of the adult *C. elegans* hermaphrodite is accompanied by the generation and subsequent deaths of an additional 131 cells (Sulston and Horvitz, *Dev. Biol.* 82:110-156 (1977); Sulston et al., *Dev. Biol.* 100:64-119 (1982)). The morphology of cells undergoing programmed cell death in *C. elegans* has been described at both the light and electron microscopic levels (Sulston and Horvitz, *Dev. Biol.* 82:100-156 (1977); Robertson and Thomson, *J. Embryol. Exp. Morph.* 67:89-100 (1982)).

Many genes that affect *C. elegans* programmed cell death have been identified (reviewed by Ellis et al., *Ann. Rev. Cell Biol.* 7:663-698 (1991)). The activities of two of these genes, *ced-3* and *ced-4*, are required for the onset of almost all *C. elegans* programmed cell deaths (Ellis and Horvitz, *Cell* 44:817-829 (1986)). When the activity of either *ced-3* or *ced-4* is eliminated, cells that would normally die instead survive and can differentiate into recognizable cell types and even function (Ellis and Horvitz, *Cell* 44:817-829 (1986); Avery and Horvitz, *Cell* 51:1071-1078 (1987); White et al., *Phil. Trans. R. Soc. Lond. B.* 331:263-271 (1991)). Genetic mosaic analyses have indicated that the *ced-3* and *ced-4* genes most likely act in a cell autonomous manner within dying cells, suggesting that the products of these genes are expressed within dying cells and either are cytotoxic molecules or control the activities of cytotoxic molecules (Yuan and Horvitz, *Dev. Biol.* 138:33-41 (1990)).

Summary of the Invention

This invention relates to genes shown to be essential for programmed cell death, referred to herein as cell death genes, to their encoded products (RNA and

polypeptides), and to antibodies directed against the encoded polypeptides. Methods and probes for identifying and screening for other cell death genes, including those of vertebrates as well as  
5 invertebrates, and possibly, microbes and plants, are described. Agents which mimic or affect the activity of cell death genes and methods for identifying these agents are also described. Bioassays which detect the activity of cell death genes and which are useful for  
10 identifying cell death genes, for testing the effect of mutations in cell death genes, and for identifying agents which mimic or affect the activity of cell death genes are also provided. This invention further relates to methods for altering (increasing or  
15 decreasing) the activity of the cell death genes or their encoded products in cells and, thus, for altering the proliferative capacity or longevity of a cell population or organism.

Specifically, the *ced-3* and *ced-4* genes of the  
20 nematode *C. elegans* have been identified, sequenced, and characterized. These genes have been shown to be required for almost all the programmed cell deaths which occur during development in *C. elegans*. Thus, two cell death genes and their encoded products (RNA,  
25 polypeptide) are now available for a variety of uses.

As described herein, the *ced-3* and *ced-4* genes can be used to identify structurally related genes from a variety of sources. Some of these related genes are likely to also function as cell death genes.  
30 Structural comparison of related cell death genes, as well as mutational analysis, can provide insights into functionally important regions or features of cell death genes and gene products. This information is useful in the design of agents which mimic or which  
35 alter the activity of cell death genes.

This invention further provides methods and agents for altering (increasing or decreasing) the occurrence of cell death in a cell population or organism. Methods and agents, described herein, which decrease cell death are potentially useful for treatment (therapeutic and preventive) of disorders and conditions characterized by cell deaths, including myocardial infarction, stroke, traumatic brain injury, degenerative diseases (e.g., Huntington's disease, amyotrophic lateral sclerosis, Alzheimer's disease, Parkinson's disease, and Duchenne's muscular dystrophy), viral and other types of pathogenic infection (e.g., human immunodeficiency virus, HIV), aging and hair loss. Methods and agents which increase cell death are also provided and are potentially useful for reducing the proliferation or size of cell populations, such as cancerous cells, cells infected with viruses (e.g., HIV) or other infectious agents, cells which produce autoreactive antibodies and hair follicle cells. Such methods and agents may also be used to incapacitate or kill undesired organisms, such as pests, parasites, and recombinant organisms.

Brief Description of the Drawings

Figure 1 shows the genomic organization and nucleotide sequence (Seq. ID #1) of ced-4 and deduced amino acid sequence (Seq. ID #2). The genomic sequence of the ced-4 region was obtained from plasmid C10D8-5, which rescues the ced-4 mutant phenotype. Two likely transcriptional start sites are marked with downward arrows. The start of the cDNA is marked with a solid arrowhead. The positions of eight ced-4 mutations are indicated by upward arrows. Numbers on the sides indicate nucleotide positions, beginning at the start

of C10D8-5. Numbers under the amino acid sequence indicate codon positions. Vertical lines between nucleotides indicate splice junctions.

Figure 2 shows the genomic structure of the ced-4 gene and positions of ced-4 mutations. The sizes of exons and introns are indicated in base pairs (bp). The downward arrows indicate the positions of the Tc4 insertion in the ced-4(n1416) mutant and of eight EMS-induced mutations of ced-4. The arrow pointing right indicates the direction of transcription. The solid arrowhead indicates the translation initiation site. The open arrowhead indicates the ochre termination codon.

Figure 3 shows the sequence similarities between the Ced-4 protein and some calcium-binding proteins. The consensus sequence of the calcium-binding loop is shown at the top. The positions indicated by X, Y, Z, -X, and -Z correspond to vertices of an octahedron. The numbers above the X, Y, Z, -X and -Z correspond to the positions of the residues within the 29 amino acid EF-hand sequence. Amino acids are indicated by the single letter code. O, amino acid with an oxygen-containing side chain. \*, non-conserved amino acid. Positions Y, Z and -X can be any amino acid with oxygen-containing side chains. Position X is usually aspartic acid, and position -Z is usually glutamic acid. Conserved amino acids are shown in bold-face. Deviations from the EF-hand consensus sequence are underlined.

Figure 4 shows the nucleotide sequence (Seq. ID #18) of ced-3 and deduced amino acid sequence (Seq. ID #19). The genomic sequence of the ced-3 region was obtained from plasmid pJ107, which rescues the ced-3 mutant phenotype. The likely translation initiation site is indicated by a solid arrowhead. The SL1 splice

DEPARTMENT OF BIOLOGY  
UNIVERSITY OF TORONTO

acceptor of the RNA is boxed. The positions of 12 ced-3 mutations are indicated. Repetitive elements in the introns are indicated as arrows above the relevant sequence. Numbers on the sides indicate nucleotide 5 positions, beginning with the start of pJ107. Numbers under the amino acid sequence indicate codon positions.

Figure 5A shows the genomic structure of the ced-3 gene and the location of the mutations. The sizes of the introns and exons are given in bp. The downward 10 arrows indicate the positions of 12 EMS-induced mutations of ced-3. The arrow pointing right indicates the direction of transcription. The solid arrowhead indicates the translation initiation site. The open arrowhead indicates the termination codon.

15 Figure 5B shows the locations of the mutations relative to the exons (numbered 1-8) and the encoded serine-rich region.

Figure 6 is a Kyte-Doolittle hydrophobicity plot of the Ced-3 protein.

20 Figure 7 shows a comparison of the Ced-3 proteins of *C. elegans* (line 1) and related nematodes, *C. briggsae* (line 2) and *C. vulgaris* (line 3). The conserved amino acids are indicated by ". ". Gaps inserted in the sequence for the purpose of alignment 25 are indicated by "\_".

Figure 8 shows a restriction site map of the ced-4 region and the relative positions of plasmid C10D8-5, plasmid insert pn1416, and three transcripts encoded by the region.

30 Figure 9 shows physical and genetic maps of the ced-3 region on chromosome IV.

Figure 10 summarizes experiments to localize ced-3 within C48D1. Restriction sites of plasmid C48D1 and subclone plasmids are shown. ced-3 activity was scored 35 as the number of cell corpses in the head of L1 young

animals. ++, the number of cell corpses above 10. +, the number of cell corpses below 10 but above 2. -, the number of cell corpses below 2.

Detailed Description of the Invention

5       The *ced-3* and *ced-4* genes of *C. elegans* have been shown to be required for almost all programmed cell deaths in *C. elegans* development (Ellis and Horvitz, *Cell* 44:817-829 (1986)). The present work describes the cloning, sequencing and characterization of these  
10 genes. As a result of this work, two genes whose activities are required for cell death, referred to herein as cell death genes, and their encoded products (RNA, polypeptide) are available for a variety of uses. Described below are the cloning and characterization of  
15 the *C. elegans ced-4* and *ced-3* genes, methods and probes for identifying structurally related genes, methods for identifying cell death genes from a variety of organisms, methods for identifying agents which mimic or which affect the activity of cell death genes,  
20 and methods and agents for altering cell death activity and thus, for altering the occurrence of cell death in a cell population or organism.

The activity of a cell death gene is intended to include the activity of the gene itself and of the  
25 encoded products of the gene. Thus, agents and mutations which affect the activity of a gene include those which affect the expression as well as the function of the encoded RNA and protein. The agents may interact with the gene or with the RNA or protein  
30 encoded by the gene, or may exert their effect more indirectly.

The ced-4 Gene

The cloning, sequencing and characterization of the *C. elegans* ced-4 gene are described in Example 1. Genomic clones were obtained from a ced-4 mutant allele.

5 generated by transposon tagging. A subclone containing as little as 4.4 kb of wild-type genomic DNA was shown to complement the ced-4 mutant phenotype (see Table 1; tables are located at the end of the Detailed Description).

10 A 2.2 kb mRNA was identified as the ced-4 transcript. The transcript was shown to be present at normal levels in a ced-3 mutant, suggesting that ced-3 is not a transcriptional regulator of ced-4 gene expression. Furthermore, the 2.2 kb transcript was

15 shown to be expressed primarily during embryogenesis. This is consistent with the observation that 113 of the 131 programmed cell deaths in *C. elegans* are embryonic (Sulston and Horvitz, *Dev. Biol.* 82:110-156 (1977); Sulston et al., *Dev. Biol.* 100:64-119 (1983)).

20 cDNA clones were further obtained and sequenced. Analysis of the cDNA and its encoded product indicates that the putative Ced-4 protein is 549 amino acids in length (Figure 1; Seq. ID #2) and about 62,877 in relative molecular mass. The Ced-4 protein is highly

25 hydrophilic, with a predicted pI of 5.12; there are no obvious transmembrane regions. The longest hydrophobic region is a segment of 12 amino acids from residues 382 to 393.

Sequence analysis of the ced-4 genomic clone and

30 comparison with the cDNA sequence revealed that the ced-4 gene contains 7 introns with sizes ranging from 44 bp to 557 bp (Figure 2).

The nucleotide sequences of eight EMS-induced ced-4 mutations were also determined. Of the eight

35 mutations, one results in a single amino acid

substitution and the other seven appear to prevent either *ced-4* RNA splicing or completion of *Ced-4* protein synthesis (Figure 2 and Table 2). These seven mutations establish the null phenotype of the *ced-4* gene, confirming that *ced-4* function is not essential for viability.

Two regions of the inferred *Ced-4* protein have sequence similarity to known calcium-binding domains (Kretsinger, *Cold Spring Harbor Symp. Quant. Biol.* 10 52:499-510 (1987)), suggesting that *Ced-4* activity and hence, programmed cell death may be modulated by calcium (see Figure 3 and Example 1). Calcium has been implicated as an essential mediator of cell death in other organisms under a variety of conditions. For example, extracellular calcium is required for glucocorticoid-induced thymocyte death (Cohen and Duke, *J. Immunol.* 132:38-42 (1984)), for the deaths of adult rat hepatocytes induced by certain toxins *in vitro* (Schanne et al., *Science* 206:700-702 (1979)), for agonist-induced muscle degeneration in mice (Leonard and Salpeter, *J. Cell Biol.* 82:811-819 (1979)) and for neuronal cell death caused by oxygen deprivation or excitotoxicity (Coyle et al., *Neurosci. Res. Prog. Bull.* 19:331-427 (1981); Choi, *J. Neurosci.* 7:369-379 (1987), Choi, *Trends Neurosci.* 11:465-469 (1988)). It is possible that programmed cell death is initiated during *C. elegans* development by an increase in intracellular calcium, which activates the *Ced-4* protein to become cytotoxic. On the other hand, certain cells seem to be protected against cell death by calcium (e.g., Koike et al., *Proc. Natl. Acad. Sci. USA* 86:6421-6425 (1989); Collins et al., *J. Neurosci.* 11:2582-2587 (1991)), suggesting that increases in intracellular calcium levels may inhibit the activity

of the Ced-4 protein and thereby prevent programmed cell death.

The level of the *ced-4* transcript in eggs is about 20% that of the actin 1 transcript, which is relatively abundant (Edwards and Wood, *Dev. Biol.* 97:375-390 (1983)). This level seems higher than might be expected if *ced-4* were expressed only in dying cells, since in an embryo there are usually no more than two or three cells dying at the same time. These considerations suggest that *ced-4* might be transcribed not only in dying cells but in other cells as well. Perhaps Ced-4 activity, at least during embryonic development, is regulated at a post-transcriptional level. For example, the Ced-4 protein might have to interact with other proteins or other factors (such as calcium) to cause cell death. Since the *ced-3* gene is also essential for programmed cell death in *C. elegans*, one possibility is that the activity of the Ced-4 protein is dependent upon *ced-3* function.

20

#### The *ced-3* Gene

The cloning, sequencing and characterization of the *ced-3* gene are described in Example 2. The *ced-3* gene was cloned by mapping DNA restriction fragment length polymorphisms (RFLPs) and chromosome walking. A 7.5 kb fragment of genomic DNA was shown to complement *ced-3* mutant phenotypes. A 2.8 kb transcript was further identified. The *ced-3* transcript was found to be most abundant in embryos, but was also detected in larvae and young adults, suggesting that *ced-3* is not only expressed in cells undergoing programmed cell death.

A 2.5 kb cDNA corresponding to the *ced-3* mRNA was sequenced. The genomic sequence was also determined (Figure 4; Seq. ID #18) and a comparison with the cDNA

sequence revealed that the *ced-3* gene has 8 introns which range in size from 54 to 1195 bp (Figure 5A). The four largest introns as well as sequences 5' of the start codon contain repetitive elements, some of which 5 have been previously characterized in non-coding regions of other *C. elegans* genes such as *fem-1* (Spence et al., *Cell* 60:981-990 (1990)), *lin-12* (J. Yochem, personnal communication), and *myoD* (Krause et al., *Cell* 63:907-919 (1990)). The transcriptional start site was 10 also mapped, and the *ced-3* transcript was found to be trans-spliced to a *C. elegans* splice leader, SL1.

Twelve EMS-induced *ced-3* alleles were also sequenced. Eight of the mutations are missense mutations, two are nonsense mutations, and two are 15 putative splicing mutations (Table 3). The molecular nature of these mutations, together with results of genetic and developmental analyses of nematodes homozygous for these mutations, indicate that, like *ced-4*, *ced-3* function is not essential to viability. 20 In addition, 10 out of the 12 mutations are clustered in the C-terminal region of the gene (Figure 5B), suggesting that this portion of the encoded protein may be important for activity.

The *ced-3* gene encodes a putative protein of 503 25 amino acids (Figure 4; Seq. ID #19). The protein is very hydrophilic and no significantly hydrophobic region can be found that might be a transmembrane domain (Figure 6). One region of the *ced-3* protein is very rich in serine. Sequence comparison of two 30 additional *ced-3* genes from related nematodes, *C. briggsae* and *C. vulgaris*, suggests that the exact sequence in this serine-rich region may not be important but that the serine-rich feature is (Figure 7; Seq. ID #19-21). This hypothesis is supported by 35 the analysis of *ced-3* mutations: none of 12 EMS-induced

09081729-202302

ced-3 mutations is in the serine-rich region (Figure 5B).

The conservation of the serine-rich feature among the ced-3 genes of different nematodes suggests that the serine-rich region may act in semi-specific protein-protein interactions, similar to acid blobs in transcription factors and basic residues in nuclear localization signals. In all these cases, the exact primary sequence is not important.

It is possible that the serine residues in the Ced-3 and Ced-4 proteins may be targets for a Ser/Thr kinase, and that the activity of these proteins may be regulated post-translationally by protein phosphorylation. McConkey et al. (*J. Immunol.*, 145:1227-1230 (1990)) have shown that phorbol esters, which stimulate protein kinase C, can block the death of cultured thymocytes induced by exposure to  $\text{Ca}^{++}$ , ionophores or glucocorticoids (Wyllie, *Nature* 284:555-556 (1980); Wyllie et al., *J. Path.* 142:67-77 (1984)).

It is possible that protein kinase C may inactivate certain cell death proteins by phosphorylation, and thus, inhibit cell death and promote cell proliferation. Several agents that can elevate cytosolic cAMP levels have been shown to induce thymocyte death, suggesting that protein kinase A may also play a role in mediating thymocyte death. Further evidence suggests that abnormal phosphorylation may play a role in the pathogenesis of certain cell-degenerative diseases. For example, abnormal phosphorylation of the microtubule-associated protein Tau is found in the brains of Alzheimer's disease and Down's syndrome patients (Grundke-Iqbali et al., *Proc. Natl. Acad. Sci. USA* 83:4913-4917 (1986); Flament et al., *Brain Res.* 516:15-19 (1990)). Thus, it is possible that phosphorylation may have a role in

regulating programmed cell death in *C. elegans*. This is consistent with the fairly high levels of *ced-3* and *ced-4* transcripts which suggest that transcription regulation alone may be insufficient to regulate  
5 programmed cell death.

Structurally and Functionally Related Genes

As a result of the work described herein, it is possible to identify genes which are structurally  
10 and/or functionally related to *ced-3* or *ced-4*. Such genes are expected to be found in a variety of organisms, including vertebrates (e.g., mammals and particularly humans), invertebrates (e.g., insects), microbes (e.g., yeast) and possibly plants.  
15 Structurally related genes refer herein to genes which have some structural similarity to the nucleotide sequences (genomic or cDNA) of one or both of the *ced-3* or *ced-4* genes, or whose encoded proteins have some similarity to one or both of the amino acid sequences  
20 of the Ced-3 or Ced-4 proteins. Functionally related genes refer to genes which have similar activity to that of *ced-3* and *ced-4* in that they cause cell death. Such genes can be identified by their ability to complement *ced-3* or *ced-4* mutations in bioassays, as  
25 described below.

Previous studies are consistent with the hypothesis that genes similar to the *C. elegans* *ced-3* and *ced-4* genes may be involved in the cell deaths that occur in both vertebrates and invertebrates. Some  
30 vertebrate cell deaths share certain characteristics with the programmed cell deaths in *C. elegans* that are controlled by *ced-3* and *ced-4*. For example, up to 14% of the neurons in the chick dorsal root ganglia die immediately after their births, before any signs of  
35 differentiation (Carr and Simpson, *Dev. Brain Res.*.

2:57-162 (1982)). Genes like *ced-3* and *ced-4* could well function in this class of vertebrate cell death.

In addition, genes related to *ced-3* and *ced-4* could function in many other types of vertebrate cell death processes, including those involving cells that die long after their births and those that die as a result of stress (e.g., oxygen deprivation) or disease.

Genetic mosaic analysis has suggested that the *ced-3* and *ced-4* genes act within cells that undergo programmed cell death, rather than through cell-cell interactions or diffusible factors (Yuan and Horvitz, *Dev. Biol.* 138:33-41 (1990)). Many cell deaths in vertebrates seem different in that they appear to be controlled by interactions with target tissues. For example, it is thought that a deprivation of target-derived growth factors is responsible for vertebrate neuronal cell deaths (Hamburger and Oppenheim, *Neurosci. Comment.* 1:39-55 (1982)); Thoenen et al., in: *Selective Neuronal Death*, Wiley, New York, 1987, Vol.

126, pp. 82-85). However, even this class of cell death could involve genes like *ced-3* and *ced-4*, since pathways of cell death involving similar genes and mechanisms might be triggered in a variety of ways.

Supporting this idea are several *in vitro* and *in vivo* studies which show that the deaths of vertebrate as well as invertebrate cells can be prevented by inhibitors of RNA and protein synthesis, suggesting that activation of genes is required for these cell deaths (Martin et al., *J. Cell Biol.* 106:829-844

(1988); Cohen and Duke, *J. Immunol.* 132:38-42 (1984); Oppenheim and Prevette, *Neurosci. Abstr.* 14:368 (1988); Stanisic et al., *Invest. Urol.* 16:19-22 (1978); Oppenheim et al., *Dev. Biol.* 138:104-113 (1990); Fahrbach and Truman, in: *Selective Neuronal Death*, Ciba Foundation Symposium, 1987, No. 126, pp. 65-81). It is

possible that the genes induced in these dying vertebrate and invertebrate cells are cell death genes similar to the *C. elegans* genes *ced-3* and *ced-4*.

Also supporting the hypothesis that cell death in *C. elegans* is mechanistically similar to cell death in vertebrates is the observation that the protein product of the *C. elegans* gene *ced-9* is similar in sequence to the human protein *Bcl-2*. *ced-9* has been shown to prevent cells from undergoing programmed cell death during nematode development by antagonizing the activities of *ced-3* and *ced-4* (Hengartner, et al., *Nature* 356:494-499 (1992)). The *bcl-2* gene has also been implicated in protecting cells against cell death. It seems likely that the genes and proteins with which *ced-9* and *bcl-2* interact are similar as well.

Genes which are structurally related to *ced-3* or *ced-4* are likely to also act as cell death genes. Structurally related genes can be identified by any number of detection methods which utilize a defined nucleotide or amino acid sequence or antibodies as probes. For example, nucleic acid (DNA or RNA) containing all or part of the *ced-3* or *ced-4* gene can be used as hybridization probes or as polymerase chain reaction (PCR) primers. Degenerate oligonucleotides derived from the amino acid sequence of the *Ced-3* or *Ced-4* proteins can also be used. Nucleic acid probes can also be based on the consensus sequences of conserved regions of genes or their protein products. In addition, antibodies, both polyclonal and monoclonal, can be raised against the *Ced-3* and/or *Ced-4* proteins and used as immunoprobes to screen expression libraries of genes.

One strategy for detecting structurally related genes in other organisms is to initially probe animals which are taxonomically closely related to the source

of the probes, for example, probing other worms with a ced-3 or ced-4 probe. Closely related species are more likely to possess related genes or gene products which are detected with the probe than more distantly related organisms. Sequences conserved between ced-3 or ced-4 and these new genes can then be used to identify similar genes from less closely related species. Furthermore, these new genes provide additional sequences with which to probe the molecules of other animals, some of which may share conserved regions with the new genes or gene products but not with ced-3, ced-4, or their gene products. This strategy of using structurally related genes in taxonomically closer organisms as stepping stones to genes in more distantly related organisms can be referred to as walking along the taxonomic tree.

Groups of structurally related genes, such as those obtained by using the above-described strategy, can be referred to as gene families. Comparison of members within a gene family, or their encoded products, may indicate functionally important features of the genes or their gene products. Those features which are conserved are likely to be significant for activity. Such conserved sequences can then be used both to identify new members of the gene family and in drug design and screening. For example, as described in Example 2, genes similar to ced-3 from two other species of nematodes (*C. briggsae* and *C. vulgaris*) were identified and characterized. Serine-rich regions were found in the polypeptides encoded by all three genes. Although the sequence of the serine-rich region was not well conserved, the number of serines was conserved, suggesting that the serine-rich feature, but not the exact sequence of the serine-rich region, is significant for function.

DEPARTMENT OF DEFENSE

Functionally important regions can also be identified by mutagenesis. For example, inactivating mutations of *ced-3* were found to cluster within a region near the COOH-terminus (Figure 5B), suggesting that this region is a functionally important domain of the *Ced-3* protein. Further mutational analyses can be carried out on the *ced-3* and *ced-4* genes; mutants with novel properties, as well as other regions important for activity, may be discovered. Mutations and other alterations can be accomplished using known methods, such as *in vivo* and *in vitro* mutagenesis (see, e.g., Ausubel et al. (eds.), *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Interscience, New York).

15 Bioassays and Agents Which Affect the Activity of Cell Death Genes

This invention further provides bioassays which detect the activity of cell death genes. The bioassays can be used to identify novel cell death genes, to identify mutations which affect the activity of cell death genes, to identify genes which are functionally related to known cell death genes, such as *ced-3* or *ced-4*, to identify genes which interact with cell death genes, and to identify agents which mimic or affect the activity of cell death genes (e.g., agonists and antagonists). For example, the bioassays can be used to screen expression gene libraries for cell death genes from other organisms.

In this bioassay, genes or agents are introduced into nematodes to test their effect on cell deaths *in vivo*. Wild-type, mutant, and transgenic nematodes can be used as appropriate for the effect being tested. In one embodiment of this bioassay, transgenic nematodes are produced using a candidate cell death gene, a

mutant cell death gene, or genes from an expression library, to observe the effect of the transgene on the pattern of programmed cell deaths during development of the nematode. For example, a gene which is

5 structurally related to ced-3 can be used to produce a transgenic animal from a mutant nematode which underexpresses or expresses an inactivated ced-3 gene to see if the related gene can complement the ced-3 mutation and is thus, functionally as well as

10 structurally related to ced-3. cDNA or genomic libraries can be screened for genes having cell death activity. Genes which interact with cell death genes to enhance or suppress their activity can also be identified by this method.

15 In another embodiment of the bioassay, wild-type, mutant, or transgenic nematodes are exposed to or administered peptides and other molecules in order to identify agents that mimic, increase, or decrease the activity of a cell death gene. For example, wild-type

20 animals can be used to test agents that inactivate or antagonize the activity of ced-3 or ced-4 and hence, decrease cell deaths, or that activate or enhance ced-3 or ced-4 activity and increase cell deaths. Mutant animals in which ced-3 or ced-4 is inactivated can be

25 used to identify agents or genes which mimic ced-3 or ced-4 in causing cell deaths. Mutant animals in which ced-3 or ced-4 is overexpressed or constitutively activated can similarly be used to identify agents that prevent ced-3 or ced-4 from causing cell death.

30 Transgenic animals in which a wild-type or mutant form of an exogenous cell death gene causes excess cell deaths due to overexpression or hyperactivity can be used to identify agents that inactivate or inhibit the activity of the transgene. Similarly, transgenic

35 animals in which a wild-type or mutant form of an

exogenous cell death gene is underexpressed or inactive can be used to identify agents that activate or increase its activity. Test molecules can be introduced into nematodes by microinjection, diffusion,  
5 ingestion, shooting with a particle gun, or other method.

Mutated cell death genes with novel properties may be identified by the above bioassay. For example, constitutively activated or hyperactive cell death  
10 genes may be isolated which may be useful as agents to increase cell deaths. Mutations may also produce genes which do not cause cell death but which antagonize the activity of the wild-type gene.

Agents can be obtained from traditional sources,  
15 such as extracts (e.g., bacterial, fungal or plant) and compound libraries, or by newer methods of rationale drug design. Information on functionally important regions of the genes or gene products, gained by sequence and/or mutational analysis, as described  
20 above, may provide a basis for drug design. The activity of the agents can be verified both by *in vivo* bioassays using nematodes which express various forms of *ced-3*, *ced-4*, or related genes, as described above, and by *in vitro* systems, in which the genes are  
25 expressed in cultured cells, or in which isolated or synthetic gene products are tested directly in biochemical experiments. The agents may include all or portions of the *ced-3*, *ced-4*, or related genes, mutated genes, and all or portions of the gene products (RNA,  
30 including antisense RNA, and protein), as well as nucleic acid or protein derivatives, such as oligonucleotides and peptides, peptide and non-peptide mimetics, and agonists and antagonists which affect the activity or expression of the cell death genes. The  
35 agents can also be portions or derivatives of genes or

gene products which are not cell death genes but which regulate the expression of, interact with, or otherwise affect the function of cell death genes or gene products.

5 Uses of the Invention

Using the above-described probes and bioassays, the identification and expression of *ced-3*, *ced-4* or related cell death genes in cultured cells, tissues, and whole organisms can be studied to gain insights 10 into their role in development and pathology in various organisms. For example, the detection of abnormalities in the sequence, expression, or activity of a cell death gene or gene product may provide a useful diagnostic for diseases involving cell deaths.

15 This invention further provides means of altering or controlling the activity of a cell death gene in a cell, and, thus, affecting the occurrence of cell death. Activity of the cell death gene can be altered to either increase or decrease cell deaths in a 20 population of cells and, thus, affect the proliferative capacity or longevity of a cell population, organ, or entire organism.

Agents which act as inactivators or antagonists of the activity of *ced-3*, *ced-4*, or other cell death genes 25 can be used to prevent or decrease cell deaths. Such agents are useful for treating (i.e., for both preventive and therapeutic purposes) disorders and conditions characterized by cell deaths, including neural and muscular degenerative diseases, stroke, 30 traumatic brain injury, myocardial infarction, viral (e.g., HIV) and other types of pathogenic infections, as well as cell death associated with normal aging and hair loss. The agent can be delivered to the affected cells by various methods appropriate for the cells or

organs being treated, including gene therapy. For example, anti-sense RNA encoded by all or a part of a cell death gene which is complementary to the mRNA can be delivered to a population of cells by an appropriate vector, such as a retroviral or adenoviral vector, or an antagonist of cell death activity can be infused into a wound area to limit tissue damage.

Methods and agents which cause or increase cell deaths are also useful, for example, for treating disorders characterized by an abnormally low rate or number of cell deaths or by excessive cell growth, such as neoplastic and other cancerous growth. Such methods and agents are also useful for controlling or eliminating cell populations, such as cells infected with viruses (e.g., HIV) or other infectious agents, cells producing autoreactive antibodies, and hair follicle cells. In addition, methods and agents which increase cell death can be used to kill or incapacitate undesired organisms, such as pests, parasites and genetically engineered organisms. All or portions of *ced-3*, *ced-4*, or related cell death genes, active mutant genes, their encoded products, agents which mimic the activity of cell death genes, and activators and agonists of cell death genes can be used for this purpose.

For example, cell death genes can be used to kill cells infected with the human immunodeficiency virus (HIV), and thus, prevent or limit HIV infection in an individual. A recombinant gene can be constructed, in which a cell death gene is under the control of a viral promoter which is specifically activated by a viral protein; the recombinant gene is introduced into HIV infected cells. HIV-infected cells containing the viral activator protein would express the cell death

gene product and be killed, and uninfected cells would be unaffected.

Alternatively, an antagonist of *ced-3* or *ced-4* activity (such as antisense RNA) can be expressed under the control of a viral-specific promoter and in this way, be used to prevent the cell death associated with viral (e.g., HIV) infection.

In another example, cell death genes can be used as suicide genes for biological containment purposes.

10 Genetic engineering of suicide genes into recombinant organisms has been reported in bacteria (*Genetic Engineering News*, Nov. 1991, p. 13): suicide genes were engineered to be expressed simultaneously with the desired recombinant gene product so that the

15 recombinant bacteria die upon completion of their task. The present invention provides suicide genes which are useful in a variety of organisms in addition to bacteria, for example in insects, fungi, and transgenic rodents. Suicide genes can be constructed by placing

20 the coding sequence of an exogenous cell death gene or an agonist of an endogenous cell death gene of the organism in an expression vector suitable for the organism.

In addition, agents which increase cell death are

25 useful as pesticides (e.g., anthelmintics, nematicides). For example, many nematodes are human, animal, or plant parasites. *ced-3*, *ced-4*, or other nematode cell death genes, their gene products, mimetics, and agonists can be used to reduce the

30 nematode population in an area, as well as to treat individuals already infected with the parasite or protect individuals from infection. A transgenic plant or animal carrying a constitutively activated *ced-3* gene, *ced-4* gene, or other cell death gene specific to

nematodes can be protected from nematode infection in this way.

The subject invention will now be illustrated by the following examples, which are not intended to be

5 limiting in any way.<sup>2</sup>

EXAMPLE 1

CLONING, SEQUENCING AND CHARACTERIZATION OF  
THE CED-4 GENE

MATERIALS AND METHODS

10 General Methods and Strains

Techniques used for the culturing of *C. elegans* were essentially as described by Brenner (*Genetics* 77:71-94 (1974)). All strains were grown at 20°C. DNA was prepared from worms grown on Petri dishes containing agarose seeded with *E. coli* strain HB101. RNA was prepared from mass cultures grown in liquid. Usually, the bacterial pellet from a 2 L overnight culture of *E. coli* HB101 grown in superbroth (12 g Bacto-tryptone, 24 g yeast extract, 8 ml 50% glycerol, 20 900 ml H<sub>2</sub>O; after autoclaving, 100 ml 0.17 M KH<sub>2</sub>HPO<sub>4</sub> and 0.72 K<sub>2</sub>HPO<sub>4</sub> were added) was resuspended in 500 ml S basal medium (Brenner, 1974 *supra*), and worms were added from one or two 10 cm Petri dishes in which the bacterial lawns had just been consumed. Worms were 25 harvested about 4-5 days later by centrifugation and washed in M9 buffer (Brenner, 1974 *supra*). The yield was about 5-10 ml of packed worms.

Nomarski differential interference contrast microscopy was used to examine individual cells in 30 living nematodes (Sulston and Horvitz, *Dev. Biol.* 82:110-156 (1977)). Methods for scoring the Ced phenotype of ced-1, ced-4 and ced-1; ced-4 double

mutants have been described by Ellis and Horvitz, (*Cell* 44:817-829 (1986)) and by Yuan and Horvitz, (*Dev. Biol.* 138:33-41 (1990)).

The wild-type parent of all mutant strains used in  
5 these experiments was *C. elegans* variety Bristol strain  
N2 (Brenner, 1974 *supra*). The genetic markers used are  
listed below. These markers have been described  
(Brenner, 1974 *supra*; Hodgkin et al., in: *The Nematode Caenorhabditis elegans*, Wood and the Community of *C.*  
10 *elegans* Researchers (eds.), Cold Spring Harbor  
Laboratory, New York, 1988, pp. 491-584; Finney et al.,  
*Cell* 55:757-769 (1988)). The strain TR679 carries the  
mutator *mut-2(r459)* (Collins et al., *Nature* 328:726-728  
(1987)). The *ced-4* alleles *n1894*, *n1920*, *n1947*, *n1948*,  
15 *n2247*, and *n2273* were characterized in the present  
work. Genetic nomenclature follows the standard system  
for *C. elegans* (Horvitz et al., *Mol. Gen. Genet.*  
175:129-133 (1979)):

LG I:      *ced-1(e1735)*, *unc-54(r323)*  
20      LG III:    *unc-86(n1351)*, *ced-4(n1162*, *n1416*,  
             *n1894*, *n1920*, *n1947*, *n1948*, *n2247*,  
             *n2273*, *n1416 n1712*, *n1416 n1713*),  
             *unc-79(e1068)*, *dpy-17(e164)*  
LG IV:     *unc-31(e928)*, *ced-3(n717)*  
25      LG V:     *egl-1(n986)*, *unc-76(e911)*

#### Genomic Libraries

A 4-6 kb size-selected phage library was  
constructed from *ced-4(n1416)* DNA as follows. Genomic  
DNA was digested with *Hind*III and run on a low-melting  
30 agarose gel. DNA migrating within the 4-6 kb size  
range was excised, and the low-melting agarose was  
removed by phenol extraction and precipitation  
(Maniatis et al., *Molecular Cloning: A Laboratory  
Manual*, Cold Spring Harbor Laboratory (1983)). These

DNA fragments were ligated to HindIII-digested DNA from phage  $\lambda$ NM1149 (Murray, *Phage Lambda and Molecular Cloning*, Cold Spring Harbor Laboratory, 1983, pp. 395-432). The product DNA was packaged with packaging extract from Promega. This library had a total of 140,000 plaque-forming units (pfu), of which 70% were recombinants, as estimated from the ratio of pfu on bacteria C600hfl and C600.

The phage genomic library (provided by J. Sulston) was prepared by partial digestion of wild-type *C. elegans* genomic DNA with Sau3A and cloning into the BamHI site of phage vector  $\lambda$ 2001 (Karn et al., *Gene* 32:217-224 (1984)).

#### Tc4 Probe

The Tc4 probe used for cloning the ced-4 gene and for Southern blots was Tc4-n1351, which contains a Tc4 element isolated from an unc-86(n1351) mutant strain (Finney et al., *Cell* 55:757-769 (1988); Yuan et al., *Proc. Natl. Acad. Sci. USA* 88:3334-3338 (1991)). DNA was labelled with  $^{32}$ P using either the nick-translation procedure described by Maniatis et al. (1983 *supra*) or the oligo-labelling procedure described by Feinberg and Vogelstein (*Anal. Biochem.*, 132:6-13 (1983)).

#### RNA Preparation, Northern Blot and Primer Extension

Total *C. elegans* RNA was extracted using guanidine isothiocyanate (Kim and Horvitz, *Genes & Dev.* 4:357-371 (1990)). Poly(A)<sup>+</sup> RNA was selected from total RNA by a poly(dT)-column (Maniatis et al., 1983 *supra*). To prepare stage-synchronized animals, eggs were obtained from gravid *C. elegans* adults grown at 20°C in liquid culture. A 5 - 10 ml sample of animals was treated with 50 ml of NaOCl/NaOH solution (10 ml NaOCl, 1 g NaOH, 40 ml H<sub>2</sub>O) for about 10 minutes with vortexing

until the adults were dissolved. Eggs were centrifuged and washed three times with M9 buffer. Isolated eggs were allowed to hatch in S basal medium without food for 14 hours at 20°C with shaking. L1 larvae were 5 collected by low-speed centrifugation after growth on *E. coli* HB101 for 2 hours, L2 larvae after 12 hours, L3 larvae after 24 hours, L4 larvae after 36 hours and adults after 48 hours. Northern blot analysis using DNA probes was performed essentially as described by 10 Meyer and Casson (*Genetics* 106:29-44 (1986)), except that RNA was transferred from the gel to the Gene Screen filter (DuPont, Wilmington, DE) by capillary action.

Quantitation of *ced-4* expression during embryonic 15 development was done by hybridizing two duplicate northern blots with *ced-4* cDNA clone SK2-2 and with a genomic DNA clone for the actin 1 gene, pW-16-210, which hybridizes to the 3' untranslated region of the actin 1 transcript (Krause and Hirsh, in: *Molecular 20 Biology of the Cytoskeleton*, Borisy et al. (eds.), Cold Spring Harbor Laboratory, 1984, pp. 287-292). The two probes were of the same specific activity ( $4 \times 10^8$  counts/minute/ $\mu$ g). The emission of  $\beta$  particles from the *ced-4* and actin 1 bands was counted using a  $\beta$  25 counter (Betagen, Waltham, MA). The readings were 7.7 counts/minute for the actin 1 band and 1.4 counts/ minute for the *ced-4* band.

The primer extension protocol was that of Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, 2nd 30 edition, Cold Spring Harbor Laboratory, 1989, pp. 7.79- 7.83), using the primer ATTGGCGATCCTCTCGA (Seq. ID #22). To define the lengths of the reaction products, a sequencing reaction using this primer and C10D8-5 as template was run adjacent to the product of the primer 35 extension reaction in the sequencing gel.

Direction of Transcription

The direction of transcription was determined by hybridizing northern blots with single-stranded RNA probes. The Bluescribe plasmid containing the insert pn1416 was linearized by digestion with either BamHI or HindIII, which cleaved at one or the other end of the insert. The linearized product was transcribed using T3 or T7 RNA polymerase, respectively, generating RNA from each strand. These RNA products were used to probe Northern blots according to a protocol developed by Z. Liu and V. Ambros: Filters were prehybridized in 50% formamide, 50 mM sodium phosphate (pH 6.5), 5 x SSC, 8 X Denhardt's, 0.5% SDS, 250 µg/ml salmon sperm DNA and then hybridized with probe at 55°C and washed in 4 x SSC, 0.1% SDS at 60°C 3 times for 20 minutes each and then in 2 x SSC, 0.1% SDS once at 60°C for 20 minutes. Northern blot experiments showed that the single-stranded RNA probe transcribed by T3 RNA polymerase hybridized to the 2.2 kb ced-4 mRNA, while the probe made by T7 RNA polymerase did not. This result indicates that the direction of the transcription is from the BamHI site toward the HindIII site of pn1416.

Determination of DNA Sequence

For determining DNA sequences, serial deletions were made according to Henikoff (Gene 28:351-359 (1984)). DNA sequences were determined using Sequenase and protocols obtained from US Biochemicals (Cleveland, OH). The ced-4 DNA sequence was confirmed by sequencing both strands of cDNA and genomic DNA clones.

Cloning of the Cosmid Fragment C10D8-5

The cosmid C10D8 was digested with EcoRI. Two EcoRI fragments of 2.2 kb (r5) and 2.4 kb (r7), both of

which hybridized to a mixture of *ced-4* cDNA subclones SK2-1 and SK2-2, were isolated. r7, which hybridized to SK2-1, which contains the 3' half of *ced-4* cDNA clone SK2, was cloned into the EcoRI site of plasmid 5 pBSKII (Stratagene). The EcoRI site at the 3' end of r7 was deleted by digesting with *Styl*, which cut once at 0.2 kb from the 3' end of the insert, and *Sall*, which cut once in the polylinker, and then religating. The deleted r7 plasmid was linearized with EcoRI and 10 ligated with EcoRI-digested r5, which hybridized to Sk2-2, the 5' half of *ced-4* cDNA SK2. Clones were analyzed for the correct orientation of the r5 insert based on the cDNA restriction map. One such correctly oriented clone was named C10D8-5.

15 Microinjection and Transformation

The procedure for microinjecting DNA into the gonad to obtain germline transformants was basically that of Fire (EMBO J. 5:2673-2680 (1986)) with modifications introduced by J. Sulston. Cosmid DNA to 20 be injected was purified twice using CsCl-gradient centrifugation (Maniatis et al., 1983 *supra*). Plasmid DNA to be injected was prepared by alkaline minipreps (Maniatis et al., 1983 *supra*). DNA was treated with RNAase A (37°C, 30 minutes) and then with proteinase K 25 (55°C, 30 minutes), extracted with phenol and then chloroform, precipitated twice (first in 0.3 M sodium acetate and then in 0.1 M potassium acetate, pH 7.2), and resuspended in 5 ul of injection buffer (Fire, 1986 *supra*). DNA concentrations used for injection were 30 0.1-1.0 mg/ml.

All transformation experiments used a *ced-1*; *ced-4*(n1162); *unc-31* strain as the recipient. The expression of the Ced-4 phenotype was quantified by counting the number of cell corpses in the heads of

young L1 animals. The cosmid C10D8 or plasmid subclones of C10D8 were mixed with cosmid C14G10, which contains the wild-type *unc-31(+)* gene, at a ratio of 2:1 or 3:1 to increase the likelihood that a  
5 phenotypically non-Unc transformant would contain the cosmid or plasmid being tested. Generally, 20-30 animals were injected in one experiment. Non-Unc F1 progeny of injected animals were isolated three to four days later. About 1/2 to 1/3 of the non-Unc progeny  
10 transmitted the non-Unc phenotype to their progeny and could be established as lines of transformants. Young L1 non-Unc progeny of such non-Unc transformants were examined using Nomarski optics to determine the number of cell corpses present in the heads.

15 Ced-4 Fusion Protein and Antibody Preparation

To express a Ced-4 fusion protein in *E. coli*, a clone containing both the 5' and 3' halves of the *ced-4* cDNA (SK2-2 and SK2-1) in the expression vector pET-5a (Rosenberg et al., *Gene* 56:125-135 (1987)) was  
20 constructed. The fusion protein expressed by this vector was expected to include 11 amino acids of phage T7 gene 10 protein, 5 amino acids of linker and the 546 amino acids encoded by *ced-4* cDNA SK2. The pJ76 plasmid, which encodes this fusion protein, was  
25 transformed into bacterial strain BL21. *ced-4* fusion protein was produced by this transformed strain, as expected, and subjected to electrophoresis on a polyacrylamide gel. A band, with mobility equivalent to about  $64 \times 10^3$  Mr, specific to the transformed  
30 strain was excised and used to immunize three rabbits. Sera from all three rabbits tested positive on western blots (Towbin et al., *Proc. Natl. Acad. Sci. USA* 76:4350-4354 (1979)). These sera were purified using immunoblots (Harlow and Lane, *Antibodies: A*

Laboratory Manual, Cold Spring Harbor Laboratory,  
1988).

## RESULTS

### Cloning of the ced-4 Gene by Transposon Tagging

5       The *ced-4* allele *n1416* in the *C. elegans* strain TR679 was isolated, which carries the mutator *mut-2(r459)* and shows an elevated frequency of transposition elements (Collins et al., *Nature* 328:726-728 (1987); Yuan et al., *Proc. Natl. Acad. Sci. USA* 88:3334-3338 (1991)). The *ced-4(n1416)* mutation is closely linked to a newly transposed copy of the *C. elegans* transposon Tc4 (Yuan et al., 1991 *supra*). Using Tc4 as a probe, this novel Tc4 element and its flanking region was cloned as a 5 kb HindIII fragment from a 4-6 kb size-selected *ced-4(n1416)* genomic phage library. A 3 kb adjacent to this Tc4 element was isolated by digesting the 5 kb HindIII fragment with *BamHI*. This 3 kb fragment, called pn1416, was cloned into the Bluescribe M13+ plasmid vector (Stratagene).

10      When used as a probe on Southern blots, pn1416 hybridized to a 3.4 kb HindIII fragment in DNA of wild-type (strain N2) and two non-Ced revertants of *ced-4(n1416)*, *ced-4(n1416 n1712)* and *ced-4(n1416 n1713)* (Yuan and Horvitz, *Dev. Biol.* 138:33-41 (1990)), and a 5 kb HindIII fragment in *ced-4(n1416)* animals. The hybridizing band in *ced-4(n1416)* DNA is 1.6 kb larger than that of the wild-type or the revertants, indicating that an insertion of this size is present in the *ced-4(n1416)* strain and is deleted in both revertants. These observations indicate that the Tc4 insertion in *ced-4(n1416)* animals is responsible for their Ced-4 mutant phenotype and suggest that pn1416 contains at least part of the *ced-4* gene.

15

20

25

30

To isolate additional genomic DNA from the region of this Tc4 insertion, pn1416 was used to probe a *C. elegans* Bristol N2 genomic DNA phage library. Five phage clones with inserts of 10 to

5 15 kb were isolated and shown to share a 3 kb *Bam*HI-*Hind*III fragment that hybridized to pn1416. These phage clones were used to identify cosmids that hybridized to them and that were members of a 600 kb contig of overlapping cosmids (Coulson et al., *Proc. Natl. Acad. Sci. USA* 83:7821-7825 (1986)). By using the phage clones as probes to hybridize to Southern blots, a cosmid C10D8 was identified as containing all regions of genomic DNA present in all five phage clones and in pn1416.

15 The ced-4 Mutant Phenotype Can Be Rescued by a 4.4 kb DNA Fragment

To identify *ced-4*(+) DNA capable of complementing the *Ced-4* mutant phenotype, the cosmid C10D8 was injected into the oocytes of *ced-4(n1162)* animals. To facilitate the identification of transgenic animals, a mutation in the *unc-31* gene, which affects locomotion, was included as a marker for co-transformation (Kim and Horvitz, *Genes & Dev.* 4:357-371 (1990)). Cosmid C14G10, which contains the wild-type allele of *unc-31* and does not have *Ced-4*-rescuing activity was coinjected with cosmid C10D8 into *ced-1(e1735)*; *unc-31(e928)*; *ced-4(n1162)* animals. The *ced-1* mutation was included to facilitate the scoring of the *ced-4* mutant phenotype (Ellis and Horvitz, *Cell* 44:817-829 (1986)). Specifically, when a cell undergoes programmed cell death in *C. elegans*, its corpse is quickly engulfed and destroyed by a neighboring cell (Robertson and Thomson, *J. Embryol. Exp. Morph.* 67:89-100 (1982); Sulston et al., *Dev. Biol.* 100:64-119

(1983)). A *ced-1* mutation prevents this engulfment, allowing the cell corpse to remain intact (Hedgecock et al., *Science* 220:1277-1280 (1983)). Thus, in a first or second stage (L1 or L2) *ced-1* mutant larva, many 5 cell corpses are present and can be easily visualized using Nomarski optics. *ced-4* mutations prevent cell death and the appearance of these corpses. Thus, suppression of the *Ced-4* mutant phenotype by a wild-type *ced-4* gene can be observed and readily quantified 10 in a *ced-1* mutant background based on an increase in the number of visible cell corpses.

From one such microinjection experiment, three non-Unc animals rescued for the *Unc-31* mutant phenotype were picked from among the F1 progeny, and from one of 15 them a line of non-Unc transformants was obtained. No true-breeding non-Unc animals could be isolated from this line: about 25% of the progeny of all non-Unc animals were Unc. Since no inviable zygotes were observed among the progeny of these non-Unc animals, 20 this transformant did not carry a recessive lethal insertion mutation. Rather, it seems likely that the injected DNA was maintained as an extrachromosomal array that was segregated to only some gametes, as has been reported previously for many other *C. elegans* 25 transgenic strains (e.g., Stinchcomb et al., *Mol. Cell Biol.* 82:110-156 (1985); Fire, *EMBO J.* 5:2673-2680 (1986); Way and Chalfie, *Cell* 54:5-16 (1988)). This putative extrachromosomal array was named *nEx1*. Young L1 progeny of *nEx1*-containing animals were examined 30 using Nomarski optics for the *Ced-4* phenotype.

Young L1 *ced-1* animals have an average of 23 cell corpses in the head, while *ced-1(e1735)*; *ced-4(n1162)* animals have an average of 0.6 cell corpses (Ellis and Horvitz, *Cell* 44:817-829 (1986)). Young L1 *ced-1*; 35 *ced-4(n1162)*; *nEx1* animals had an average of nine cell

corpses in the head. These results indicate that cosmid C10D8 restored significant, but not total, *ced-4(+) activity* in the transformants.

To delineate the *ced-4* gene within C10D8, various 5 subclones of C10D8 were injected into *ced-4* mutant animals and tested for their ability to rescue the *Ced-4* mutant phenotype (Table 1). The smallest subclone plasmid that could rescue the *ced-4* phenotype as effectively as cosmid C10D8 was a 4.4 kb fragment, 10 called C10D8-5. C10D8-5 and the *unc-31(+)*-containing cosmid C14G10 were coinjected into *ced-1; unc-31; ced-4(n1162)* animals. Two lines of non-Unc transformants were isolated. Since these animals continued to segregate Unc animals and did not produce 15 inviable zygotes, both appeared to carry extrachromosomal arrays, which were designated *nEx7* and *nEx8*. Young L1 animals from these transformant strains had an average of 11.5 cell corpses in their heads, indicating that plasmid C10D8-5 restored *ced-4(+) activity* as well as did cosmid C10D8 (Table 1).

#### Identification of a *ced-4* Transcript

Restriction sites of plasmid C10D8-5 (which can rescue the *Ced-4* phenotype) and pn1416 (which contains sequences adjacent to the Tc4 insertion site) were 25 mapped. C10D8-5 was found to overlap with 2 kb of sequence in pn1416, including the Tc4 insertion site (Figure 8).

In Northern blot experiments, both pn1416 and C10D8-5 were used to probe poly(A)<sup>+</sup> RNA populations of 30 mixed developmental stages of wild-type (strain N2), *ced-4(n1416)*, and *ced-4(n1416 n1712)* and *ced-4(n1416 n1713)* revertant animals. pn1416 hybridized to a 2.2 kb transcript and an 0.9 kb transcript in RNA from N2 animals, and a 3 kb transcript, a transcript slightly

larger than the wild-type 2.2 kb transcript, and a transcript slightly smaller than the wild-type 0.9 kb transcript in *ced-4(n1416)* animals. The 3.8 kb RNA contained Tc4 sequence (see below), suggesting that

5 this RNA resulted from the insertion of the 1.6 kb Tc4 sequence into the *ced-4* sequence encoding 2.2 kb transcript. The transcript slightly larger than the 2.2 kb wild-type transcript did not contain Tc4 sequence. This *ced-4(n1416)* RNA might have been an

10 aberrant transcript containing sequences adjacent to the *ced-4* gene: when pn1416 was used as a probe, the wild-type 2.2 kb and the slightly larger transcript in this mutant were relatively similar in intensities, whereas when *ced-4* cDNA clone SK2-1 was used as a

15 probe, this mutant transcript was not detected (see below). These observations indicate that the *ced-4(n1416)* 2.2 kb transcript contains sequences from the *ced-4* region but does not contain sequences corresponding to at least the 3' half of the *ced-4*

20 mRNA. The two revertants of *ced-4(n1416)*, *ced-4(n1416 n1712)* and *ced-4(n1416 n1713)*, contained both 2.2 kb and 0.9 kb transcripts with similar sizes to the wild-type transcripts. Thus, both the 2.2 kb and the 0.9 kb transcripts were altered in *ced-4(n1416)* animals, and

25 both were restored in the two non-Ced revertants.

To determine if any of the transcripts contains Tc4 sequence, the Northern blots were probed with Tc4-n1351, which contains the 1.6 kb Tc4 element present in the Tc4-induced mutant *unc-86(n1351)* as well

30 as 4 kb of *unc-86* sequences. Tc4-n1351 hybridized both to a 3.8 kb transcript of the Tc4-induced mutant *ced-4(n1416)* and to a 1.5 kb *unc-68* transcript in both *ced-4(n1416)* and N2 animals.

To determine whether one or both of the 2.2 kb and

35 0.9 kb transcripts are encoded by *ced-4*, subclone

C10D8-5, which rescued the *Ced-4* phenotype, was used to probe the Northern blots. C10D8-5 detected the wild-type 2.2 kb transcript, the *ced-4(n1416)* transcript slightly larger than the 2.2 kb transcript, and the 5 *ced-4(n1416)* 3.8 kb transcript. C10D8-5 did not hybridize to the 0.9 kb transcript, indicating that this transcript is unlikely to be encoded by *ced-4*. C10D8-5 also detected a 1.4 kb transcript, which was not altered by the Tc4 insertion in *ced-4(n1416)* 10 animals. Only a 470 bp EcoRI-StuI fragment at one end of C10D8-5 hybridized to this 1.4 kb RNA. Since C10D8-5 did not contain the complete coding region for this RNA, and since this RNA was unaffected in *ced-4(n1416)* animals, this 1.4 kb RNA seems unlikely to 15 be a *ced-4* transcript. The relationships among cosmid C10D8-5, pn1416 and the 0.9 kb, 1.4 kb and 2.2 kb transcripts are summarized in Figure 8.

On Northern blots probed with the *ced-4* cDNA clone SK2-1, the level of the 2.2 kb transcript showed 20 significant reduction in all three independently derived EMS-induced *ced-4* mutants examined, strongly supporting the hypothesis that this 2.2 kb transcript is a *ced-4* transcript. Total RNA from N2, *ced-4(n1162)*, *ced-4(n1416)*, *ced-4(n1894)* and 25 *ced-4(n1920)* eggs was probed with <sup>32</sup>P-labelled *ced-4* cDNA SK2-1. An actin 1 probe (Krause and Hirsh, in: *Molecular Biology of the Cytoskeleton*, Borisy et al. (eds.), Cold Spring Harbor Laboratory, 1984, pp. 287-292) was used as an internal control for the amount of 30 RNA loaded in each lane. The ratios of the intensity of the *ced-4* band to that of actin band in N2, *n1162*, *n1416* and *n1894* were 0.5, 0.17, 0 and 0.12, respectively. A Northern blot of poly(A)+ RNA from stage-synchronized animals was probed with pn1416, 35 which hybridizes both to the 2.2 kb *ced-4* transcript

and to a 0.9 kb transcript. The 0.9 kb transcript seems to be expressed mostly in eggs and adults. The presence of RNA in all lanes was confirmed by loading 1/10 of each sample on another gel and probing a

5 Northern blot from this gel using the *C. elegans* actin 1 gene (Krause and Hirsh, 1984 *supra*). That all of these distinct *ced-4* mutations cause reduced levels of a *ced-4* transcript could reflect either instability of all three mutant transcripts or a role for *ced-4* in

10 regulating its own expression.

Based upon these results, it can be concluded that the 2.2 kb RNA is a *ced-4* transcript. It is not known why the 0.9 kb RNA is also altered in *ced-4(n1416)* animals. Perhaps transcription of the 0.9 kb RNA is

15 initiated incorrectly as a consequence of the nearby Tc4 element.

*ced-4* Expression is Primarily Embryonic

A Northern blot containing RNAs from stage-synchronized animals of different developmental stages probed with pn1416 showed that the 2.2 kb *ced-4* transcript was expressed primarily during embryogenesis. This result is consistent with the observation that 113 of the 131 programmed cell deaths in the *C. elegans* hermaphrodite are embryonic (Sulston and Horvitz, *Dev. Biol.* 82:110-156 (1977); Sulston et al., *Dev. Biol.* 100:64-119 (1983)). The 2.2 kb RNA was relatively abundant during embryonic development. The 0.9 kb transcript was expressed mostly in eggs and adults. The presence of RNA in all lanes was confirmed by loading 1/10 of each sample on another gel and probing a Northern blot from this gel with the *C. elegans* actin 1 gene (Krause and Hirsh, 1984 *supra*).

The ced-4 Transcript is Present in a ced-3 Mutant

The activities of both *ced-3* and *ced-4* are required for programmed cell death (Ellis and Horvitz, *Cell* 44:817-819 (1986)). One possibility is that one of these genes positively regulates the expression of the other. For this reason, a Northern blot of wild-type strain N2 and *ced-3(n717)* poly(A)<sup>+</sup> RNA was probed with pn1416. This experiment showed that the 2.2 kb *ced-4* transcript was present at an apparently normal level in this *ced-3* mutant. Thus, the activity of the *ced-3* gene is unlikely to be necessary for the expression of the *ced-4* 2.2 kb transcript.

Identification of ced-4 cDNA Clones

To isolate cDNA clones of *ced-4*, pn1416 was used to probe a *C. elegans* cDNA phage library made from wild-type strain N2 mixed-stage RNA (Kim and Horvitz, *Genes & Dev.* 4:357-371 (1990)). Two cDNA clones were isolated. The two cDNA clones (named SK1 and SK2) hybridized to the 2.2 kb *ced-4* transcript. Both are about 1.8 kb in size, and both contain one 0.8 kb and one 1.0 kb EcoRI fragment. These EcoRI fragments were subcloned into plasmid vector Bluescribe M13+ (Stratagene). The two subclones derived from SK1 were named SK1-1 and SK1-2, and the two subclones derived from SK2 were named SK2-1 and SK2-2. The restriction maps of the SK1- and SK2-derived clones were the same. Sequence analysis of the ends of the four cDNA subclones confirmed the equivalence of the SK1 and SK2 clones, except that SK1-2 contains a poly(A) sequence of more than 50 bp at its 5' end. This poly(A) sequence is probably a cDNA cloning artifact, since SK1-2 contains the 5' half of the cDNA (see below).

The ced-4 Sequence

The DNA sequence of the SK2 1.8 kb cDNA clone was determined. This sequence includes an open reading frame encoding 546 amino acids (Figure 1; Seq. ID #2), which is consistent with the results of Northern blot analysis using single-stranded RNA probes. An ochre termination codon (TAA) is located in-frame near the 3' end, indicating that the 3' end of the 2.2 kb transcript is most likely included in this cDNA. The open reading frame extends to the 5' end of the 1.8 kb cDNA, suggesting that this cDNA might lack the 5' end of the ced-4 coding region.

A primer extension experiment was performed to determine the ced-4 transcription initiation site(s) using the primer ATTGGCGATCCTCTCGA (Seq. ID #23) and C10D8-5 as template. A major transcriptional initiation site was identified at 54 bp before (5' of) the beginning of the ced-4 cDNA SK2 and a minor initiation site at 54 bp after (3' of) the beginning of this cDNA (Figure 1). The first AUG codon after the presumptive major start site is located at 9 bp before the beginning of the cDNA (Figure 1). If this site is used to initiate protein synthesis, the Ced-4 protein would be 549 amino acids in length. The first AUG codon after the presumptive minor start site is located at 130 bp after the beginning of the cDNA. If this site is used, the Ced-4 protein would be 503 amino acids in length. Preliminary results using an anti-Ced-4 antibody raised against a Ced-4 fusion protein showed that endogenous Ced-4 protein is slightly smaller in molecular weight than a Ced-4 fusion protein of 562 amino acids expressed in *E. coli*. Thus, most Ced-4 protein is probably initiated near the start of the cDNA and is presumably 549 amino acids in length and 62,977 in relative molecular mass. The direction

of the open reading frame is consistent with the direction of transcription, as demonstrated by probing Northern blots with single-stranded RNA probes. The presumptive Ced-4 protein is highly hydrophilic, with a 5 pI of 5.12. The longest hydrophobic region is a segment of 12 amino acids from residues 382 to 393.

A Western blot of wild-type strain N2 mixed-stage, ced-4(*n1416*) mixed-stage, wild-type egg, and bacterially expressed protein (pJ76) was probed using 10 anti-Ced-4 antibody. Ced-4 fusion protein (pJ76) was made by cloning ced-4 cDNA SK2 into the T7 expression vector pET-5a (Rosenberg et al., Gene 56:125-135 (1987)), so that 546 amino acids of Ced-4 sequence were fused to 11 amino acids of T7 gene 10 protein and 5 15 amino acids of linker sequence. This Ced-4 fusion protein is similar in relative molecular mass to the endogenous Ced-4 protein, which is present in wild-type (N2) but missing in ced-4(*n1416*) animals. The proteins phosphorylase b,  $97 \times 10^3$ ; bovine serum albumin,  $66 \times 20 10^3$  (Hirayama et al., Biochem. Biophys. Res. Comm. 173:639-646 (1990)); and ovalbumin,  $43 \times 10^3$ , were used as molecular weight standards.

To confirm the DNA sequence obtained from the ced-4 cDNAs and to study the structure of the ced-4 25 gene, the sequences of the 4.4 kb cosmid subclone C10D8-5, the 3 kb insert pn1416, and the 2 kb HindIII-BamHI fragment that contains the Tc4 insertion in the ced-4(*n1416*) mutant were determined. Comparison of the ced-4 genomic and cDNA sequences revealed that the 30 ced-4 gene has seven introns of sizes ranging from 44 bp to 557 bp (Figure 2). The exon sequences of genomic clone C10D8-5 are identical to the sequences of ced-4 cDNA SK2. Comparison of the Tc4 insertion site in ced-4(*n1416*) DNA with the ced-4(+) genomic and cDNA

sequences indicated that Tc4 inserted into an exon in the *ced-4* gene in *ced-4(n1416)* animals (Figure 2).

The DNA sequences of eight EMS-induced *ced-4* alleles were also determined (Table 2). One of the 5 eight, *n1948*, is a missense mutation. Of the seven others, four create stop codons and three are predicted to affect splicing of the *ced-4* transcript. The positions of these mutations are indicated in Figure 2. These findings indicate that the phenotypes of these 10 mutants (Ellis and Horvitz, *Cell* 44:817-829 (1986)) result from a complete loss of *ced-4* gene function. These mutations establish the null phenotype of the *ced-4* gene, confirming that *ced-4* function is not essential for viability.

15 The Ced-4 Protein Has Two Regions Similar to Known Calcium-Binding Domains

By direct inspection, the sequence of the putative Ced-4 protein was compared with the consensus sequence of the calcium-binding loop of the EF-hand domain 20 (Tufty and Kretsinger, *Science* 187:161-171 (1975); Kretsinger, *Cold Spring Harbor Symp. Quant. Biol.* 52:499-510 (1987); Szebenyi and Moffat, *J. Biol. Chem.* 26:8761-8777 (1986)). Two regions of the Ced-4 protein were identified that might bind calcium (Figure 3).

25 The EF-hand is a 29 amino acid domain consisting of a helix-loop-helix region, with the loop portion (residues 10-21) coordinating calcium-binding via the side-chain oxygens of serine, threonine, asparagine, aspartic acid, glutamine or glutamic acid. These 30 residues occur at five of the vertices of an octahedron: X (position 10), Y (12), Z (14), -X (18), -Z (21). EF-hand amino acid sequences vary considerably in the residues present in the calcium-binding loop (Figure 3), and some EF-hand domains have

only one helical region (Kretsinger, 1987 *supra*). The consensus sequence is shown at the top of Figure 3.

Positions Y, Z, and -X can have any of a number of amino acids which have oxygen-containing side chains.

5 Position X is usually aspartic acid, and position -Z is usually glutamic acid.

The sequences of parvalbumins from carp muscle (Seq. ID #3; Nockolds et al., *Proc. Natl. Acad. Sci. USA* 69:581-584 (1972)), the intestinal calcium-binding protein (ICaBP) (Seq. ID #7-8; Szebenyi et al., *Nature* 294:327-332 (1981)), troponin C (Seq. ID #9-12; Collins et al., *FEBS Lett.* 36:268-272 (1973)) and calmodulin (Seq. ID #13; Zimmer et al., *J. Biol. Chem.* 263:19,370-19,383 (1988); Babu et al., *Nature* 315:37-40 (1985))  
10 show canonical EF-hands. The hake and ray parvalbumins (Seq. ID #4-5; Capony et al. *Eur. J. Biochem.* 32:97-108 (1973)); Thatcher and Pechere, *Eur. J. Biochem.* 75:121-132 (1977)), sarcoplasmic calcium-binding protein (SCBP) from the protochordate *Amphioxus* (Seq. ID #6;  
15 Takagi et al., *Biochemistry* 25: 3585-3592 (1986)), trypsinogen (Seq. ID #14; Bode and Schwager, *J. Mol. Biol.* 98:693-717 (1975)), fibrinogen (Seq. ID #15; Doolittle, *Ann. Rev. Biochem.* 53:195-229 (1984); Dang et al., *J. Biol. Chem.* 260:9713-9719 (1985)), villin  
20 (Seq. ID #16; Hesterberg and Weber, *J. Biol. Chem.* 258:365-369 (1983)) and galactose-binding protein (GBP) (Seq. ID #17; Vyas et al., *Nature* 327:635-638 (1987))  
25 show variations from the consensus sequence. GBP does not contain the helices of the EF-hand.  
The potential calcium-binding loops of sequence 1 and sequence 2 are located at amino acids 77-88 and amino acids 292-303 of the Ced-4 protein, respectively (Figure 3). In its putative calcium-binding loop, the first potential EF-hand-like sequence of the Ced-4  
30 protein has four (positions Y, Z, -X, -Z) of the five

conserved residues with oxygen-containing side chains (shown in bold), and the fifth position (X) has a tyrosine rather than an aspartic acid; tyrosine contains oxygen in its side chain. The second

5 potential EF-hand-like sequence of the Ced-4 protein has three residues (positions Z, -X, -Z) that match the consensus sequence, and amino acids with oxygen-containing side chains at the other two positions.

These observations suggest that these two regions of

10 the Ced-4 protein might bind calcium. Like the Ced-4 protein, a number of known calcium-binding proteins, such a bovine intestinal calcium-binding protein (ICaBP) (Szebenyi and Moffat, 1986 *supra*), rabbit troponin C (Collins et al., 1973 *supra*), trypsinogen

15 and villin (Doolittle, 1984 *supra*; Danget et al., 1985 *supra*) have only three or four conserved residues at these five positions (Figure 3). The EF-hand domains in ICaBP and troponin C have been shown by X-ray crystallography to bind calcium.

20 One major difference between the Ced-4 protein and the calcium-binding loop of the EF-hand consensus sequence is at position 15. Here, the two Ced-4 sequences have a histidine and a glutamic acid, respectively; whereas most ET-hand-containing proteins

25 have a glycine; this glycine has been suggested to be important for the turning of the loop (Kretsinger, 1987 *supra*). However, a histidine is present at this position in a parvalbumin and an aspartic acid is present in another parvalbumin and also in a

30 sarcoplasmic calcium-binding protein (Kretsinger, 1987 *supra*) (Figure 3). Thus, the presence of histidine or glutamic acid at position 15 does not rule out the possibility that these regions bind calcium.

The calcium-binding loop (positions 10-21) of the

35 EF-hand is thought to be preceded (positions 1-9) and

followed by alpha-helical domains (positions 22-29) (Kretsinger, 1987 *supra*). Since position 3 of Ced-4 sequence 1 and positions 26 and 28 of Ced-4 sequence 2 are prolines, these regions might not form alpha-  
5 helices. However, the known calcium-binding protein galactose-binding protein (GBP) has a calcium-binding domain similar to that of the EF-hand (Figure 3) but without the two helices; furthermore, position 29 of GBP is proline (Vyas et al., 1987 *supra*). Thus, the  
10 Ced-4 protein need not contain such alpha-helical calcium-binding domains.

Based upon these considerations, it seems likely that the Ced-4 protein binds calcium or a similar divalent cation.

15

#### EXAMPLE 2

#### CLONING, SEQUENCING, AND CHARACTERIZATION OF THE CED-3 GENE

#### MATERIALS AND METHODS

##### General Methods and Strains

20 The techniques used for the culturing of *C. elegans* were as described by Brenner (*Genetics* 77:71-94 (1974)). All strains were grown at 20°C. The wild-type parent strains were *C. elegans* variety Bristol strain N2, Bergerac strain EM1002 (Emmons et al., *Cell* 32:55-65 (1983)), *C. briggsae* and *C. vulgaris* (obtained from V. Ambros). The genetic markers used are described below. These markers have been described by Brenner (1974 *supra*), and Hodgkin et al. (In: *The Nematode Caenorhabditis elegans, Wood and the Community* of *C. elegans* Researchers (eds.), Cold Spring Harbor Laboratory, 1988, pp 491-584). Genetic nomenclature  
25  
30

follows the standard system (Horvitz et al., *Mol. Gen. Genet.* 175:129-133 (1979)).

LG I: *ced-1(e1375); unc-54(r323)*  
LG VI: *unc-31(e928), unc-30(e191), ced-3(n717, n718,*  
5 *n1040, n1129, n1163, n1164, n1165, n1286,*  
*n1949, n2426, n2430, n2433), unc-26(e205),*  
*dpy-4(e1166)*  
LG V: *egl-1(n986); unc-76(e911)*  
LG X: *dpy-3(e27)*

10 Isolation of Additional Alleles of *ced-3*

A non-complementation screen was designed to isolate new alleles of *ced-3*. Because animals heterozygous for *ced-3(n717)* in *trans* to a deficiency are viable (Ellis and Horvitz, *Cell* 44:817-829 (1986)),  
15 animals carrying a complete loss-of-function *ced-3*, allele generated by mutagenesis were expected to be viable in *trans* to *ced-3(n717)*, even if the new allele was inviable in homozygotes. Fourteen EMS mutagenized *egl-1* males were mated with *ced-3(n717) unc-26(e205);*  
20 *egl-1(n487); dpy-3(e27)* hermaphrodites. *egl-1* was used as a marker in this screen. Dominant mutations in *egl-1* cause the two hermaphrodite specific neurons, the HSNs, to undergo programmed cell death (Trent et al., *Genetics* 104:619-647 (1983)). The HSNs are required  
25 for normal egg-laying, and *egl-1(n986)* hermaphrodites, which lack HSNs, are egg-laying defective (Trent et al., 1983 *supra*). The mutant phenotype of *egl-1* is suppressed in a *ced-3; egl-1* strain because mutations in *ced-3* block programmed cell deaths. *egl-1* males  
30 were mutagenized with EMS and crossed with *ced-3(n717), unc-26(e205); egl-1(n487); dpy-3(e27)*. Most cross progeny were egg-laying defective because they were heterozygous for *ced-3* and homozygous for *egl-1*. Rare

egg-laying competent animals were picked as candidates for carrying new alleles of *ced-3*. Four such animals were isolated from about 10,000 F1 cross progeny of EMS-mutagenized animals. These new mutations were made 5 homozygous to confirm that they carried recessive mutations of *ced-3*.

Molecular Biology

Standard techniques of molecular biology were used (Maniatis et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1983).

Two cosmid libraries were used extensively in this work: a *Sau3AI* partial digest genomic library of 7000 clones in the vector pHC79 and a *Sau3AI* partial digest genomic library of 6000 clones in the vector pJB8 (Ish-15 Horowicz and Burke, *Nucleic Acids Res.* 9:2989 (1981)).

The "right" end of MMM-C1 was cloned by cutting it with HindIII and self-ligating. The "left" end of MMM-C1 was cloned by cutting it with *BglII* or *Sall* and self-ligating.

20 The "right" end of Jc8 was made by digesting Jc8 with *EcoRI* and self-ligating. The "left" end of Jc8 was made by digesting Jc8 by *Sall* and self-ligating.

*C. elegans* RNA was extracted using guanidine isothiocyanate (Kim and Horvitz, *Genes & Dev.* 4:357-371 (1990)). Poly(A)<sup>+</sup> RNA was selected from total RNA by a poly(dT) column (Maniatis et al., 1983 *supra*). To prepare stage-synchronized animals, worms were synchronized at different developmental stages (Meyer and Casson, *Genetics* 106:29-44 (1986)).

30 For DNA sequencing, serial deletions were made according to a procedure developed by Henikoff (*Gene* 28:351-359 (1984)). DNA sequences were determined using Sequenase and protocols obtained from US Biochemicals with minor modifications.

The Tc1 DNA probe for Southern blots was pCe2001, which contains a Bergerac Tc1 element (Emmons et al., Cell 32:55-65 (1983)). Enzymes were purchased from New England Biolabs, and radioactive nucleotides were from 5 Amersham.

Primer extension procedures followed the protocol by Robert E. Kingston (In: Current Protocols in Molecular Biology, Ausubel et al. (eds.), Greene Publishing Associates and Wiley-Interscience, New York, 10 p. 4.8.1) with minor modifications.

Polymerase chain reaction (PCR) was carried out using standard protocols supplied by the GeneAmp Kit (Perkin Elmer). The primers used for primer extension and PCR are as follows:

15 Pex2: 5' TCATCGACTTTAGATGACTAGAGAACATC 3'  
(Seq. ID #24);  
Pex1: 5' GTTGCAC TGCTTCACGATCTCCGTCTCT 3'  
(Seq. ID #25);  
SL1: 5' GTTTAATTACCCAAGTTGAG 3' (Seq. ID #26);  
20 SL2: 5' GGTTTTAACCAAGTTACTCAAG 3' (Seq. ID #27);  
Log5: 5' CCGGTGACATTGGACACTC 3' (Seq. ID #28); and  
Oligo10: 5' ACTATTCAACACTTG 3' (Seq. ID #29).

#### Germline Transformation

The procedure for microinjection basically follows 25 that of A. Fire (EMBO J. 5:2673-2680 (1986)) with modifications: Cosmid DNA was twice purified by CsCl-gradient. Miniprep DNA was used when deleted cosmids were injected. To prepare miniprep DNA, DNA from 1.5 ml overnight bacterial culture in superbroth (12 g 30 Bacto-tryptone, 24 g yeast extract, 8 ml 50% glycerol, 900 ml H<sub>2</sub>O, autoclaved; after autoclaving, 100 ml 0.17 M KH<sub>2</sub>PO<sub>4</sub> and 0.72 M KH<sub>2</sub>PO<sub>4</sub> were added) was extracted by alkaline lysis method as described in Maniatis et al.

(1983 *supra*). DNA was treated with RNase A (37°, 30 minutes) and then with protease K (55°, 30 minutes), extracted with phenol and then chloroform, precipitated twice (first in 0.3 M sodium acetate and second in 0.1 M potassium acetate, pH 7.2), and resuspended in 5  $\mu$ l injection buffer as described by A. Fire (1986 *supra*). The DNA concentration for injection is in the range of 100 ug to 1 mg per ml.

All transformation experiments used *ced-1(e1735)*; 10 *unc-31(e928) ced-3(n717)* strain. *unc-31* was used as a marker for co-transformation (Kim and Horvitz, 1990 *supra*). *ced-1* was present to facilitate scoring of the *ced-3* phenotype. The mutations in *ced-1* block the engulfment process of cell death, which makes the 15 corpses of the dead cells persist much longer than in wild-type animals (Hedgecock et al., *Science* 220:1277-1280 (1983)). The *ced-3* phenotype was scored as the number of dead cells present in the head of young L1 animals. The cosmid C10D8 or the plasmid subclones of 20 C10D8 were mixed with C14G10 (*unc-31(+)*-containing) at a ratio of 2:1 or 3:1 to increase the chances that a Unc-31(+) transformant would contain the cosmid or plasmid being tested as well. Usually, 20-30 animals were injected in one experiment. Non-Unc F1 progeny of 25 the injected animal were isolated three to four days later. About 1/2 to 1/3 of the non-Unc progeny transmitted the non-Unc phenotype to F2 progeny and established a transformant line. The young L1 progeny of such non-Unc transformant were checked for the 30 number of dead cells present in the head using Nomarski optics.

## RESULTS

### Isolation of Additional ced-3 Alleles

All of the *ced-3* alleles that existed previously were isolated in screens designed to detect viable mutants displaying the *Ced* phenotype (Ellis and Horvitz, *Cell* 44:817-829 (1986)). Such screens may have systematically missed any class of *ced-3* mutations that is inviable as homozygotes. For this reason, a scheme was designed that could isolate recessive lethal alleles of *ced-3*. Four new alleles of *ced-3* (*n1163*, *n1164*, *n1165*, *n1286*) were isolated in this way. Since new alleles were isolated at a frequency of about 1 in 2500, close to the frequency expected for the generation of null mutations by EMS in an average *C.* elegans gene (Brenner, *Genetics* 77:71-94 (1974); Greenwald and Horvitz, *Genetics* 96:147-160 (1980)), and all four alleles are homozygous viable, it was concluded that the null allele of *ced-3* is viable.

### Mapping RFLPs near ced-3

*Tc1* is a *C. elegans* transposable element that is thought to be immobile in the common laboratory Bristol strain and in the Bergerac strain (Emmons et al., *Cell* 32:55-65 (1983)). In the Bristol strain, there are 30 copies of *Tc1*, while in the Bergerac strain, there are more than 400 copies of *Tc1* (Emmons et al., 1983 *supra*; Finney, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1987). Because the size of the *C. elegans* genome is small (haploid genome size  $8 \times 10^7$  bp) (Sulston and Brenner, *Genetics* 77:95-104 (1976)), a polymorphism due to *Tc1* between the Bristol and Bergerac strains would be expected to occur about once every 200 kb. Restriction fragment

length polymorphisms (RFLPs) can be used as genetic markers and mapped in a manner identical to conventional mutant phenotypes. A general scheme has been designed to map Tc1 elements that are dimorphic

5 between the Bristol and Bergerac strains near any gene of interest (Ruvkun et al., *Genetics*, 121:501-516 (1989)). Once tight linkage of a particular Tc1 to a gene of interest has been established, that Tc1 can be cloned and used to initiate chromosome walking.

10 A 5.1 kb Bristol-specific Tc1 EcoRI fragment was tentatively identified as containing the Tc1 closest to ced-3. This Tc1 fragment was cloned using cosmids from a set of Tc1-containing *C. elegans* Bristol genomic DNA fragments. DNA was prepared from 46 such Tc1-  
15 containing cosmids, and this DNA was screened using Southern blots to identify the cosmids that contain a 5.1 kb EcoRI Tc1-containing fragment. Two such cosmids were identified: MMM-C1 and MMM-C9. The 5.1 kb EcoRI fragment was subcloned from MMM-C1 into pUC13  
20 (Promega). Since both ends of Tc1 contain an EcoRV site (Rosenzweig et al., *Nucleic Acids Res.* 11:4201-4209 (1983)), EcoRV was used to remove Tc1 from the 5.1 kb EcoRI fragment, generating a plasmid that contains only the unique flanking region of this Tc1-containing  
25 fragment. This plasmid was then used to map the specific Tc1 without the interference of other Tc1 elements.

unc-30(e191) ced-3(n717) dpy-4(e1166)/+++ males were crossed with Bergerac (EM1002) hermaphrodites, and  
30 Unc non-Dpy or Dpy non-Unc recombinants were picked from among the F2 progeny. The recombinants were allowed to self-fertilize, and strains that were homozygous for either unc-30(e191) dpy-4(Bergerac) or unc-30(Bergerac) dpy-4(e1166) were isolated. After  
35 identifying the ced genotypes of these recombinant

strains, DNA was prepared from these strains. A Southern blot of DNA from these recombinants was probed with the flanking sequence of the 5.1 kb EcoRI Tc1 fragment. This probe detects a 5.1 kb fragment in

5 Bristol N2 and a 3.4 kb fragment in Bergerac. Five out of five *unc-30 ced-3 dpy(+Berg)* recombinants, and one of one *unc-30(+Berg) ced-3 dpy-4* recombinants showed the Bristol pattern. Nine of ten *unc-30(+Berg) dpy-4* recombinants showed the Bergerac pattern. Only one

10 recombinant of *unc-30(+Berg) dpy-4* resulted from a cross-over between *ced-3* and the 5.1 kb Tc1 element. The genetic distance between *ced-3* and *dpy-4* is 2 map units ( $\mu$ ). Thus, this Tc1 element is located 0.1  $\mu$  on the right side of *ced-3*.

15 Cosmids MMM-C1 and MMM-C9 were used to test whether any previously mapped genomic DNA cosmids overlapped with these two cosmids. A contig of overlapping cosmids was identified that extended the cloned region near *ced-3* in one direction.

20 To orient MMM-C1 with respect to this contig, both ends of MMM-C1 were subcloned and these subclones were used to probe the nearest neighboring cosmid C48D1. The "right" end of MMM-C1 does not hybridize to C48D1, while the "left" end does. Therefore, the "right" end

25 of MMM-C1 extends further away from the contig. To extend this contig, the "right" end of MMM-C1 was used to probe the filters of two cosmid libraries (Coulson et al., Proc. Natl. Acad. Sci. USA 83:7821-7825 (1986)). One clone, Jc8, was found to extend MMM-C1 in

30 the opposite direction of the contig.

RFLPs between the Bergerac and Bristol strains were used to orient the contig with respect to the genetic map. Bristol (N2) and Bergerac (EM1002) DNA was digested with various restriction enzymes and

35 probed with different cosmids to look for RFLPs. Once

such an RFLP was found, DNA from recombinants of the Bristol and Bergerac strains between *ced-3* and *unc-26*, and between *unc-30* and *ced-3* was used to determine the position of the RFLP with respect to *ced-3*.

5       The "right" end of Jc8, which represents one end of the contig, detects an RFLP (*nP33*) when N2 and EM1002 DNA was digested with *HindIII*. A Southern blot of DNA from recombinants between three *ced-3(+Berg)* *unc-26* was probed with the "right" end of Jc8. Three  
10      of three *+Berg unc-26* recombinants showed the Bristol pattern, while two of two *ced-3 unc-26(+Berg)* recombinants showed the Bergerac pattern. Thus, *nP33* mapped very close or to the right side of *unc-26*.

15      The "left" end of Jc8 also detects a *HindIII* RFLP (*nP34*). The same Southern blot was reprobed with the Jc8 "left" end. Two of the two *ced-3 unc-26(+Berg)* recombinants and two of the three *ced-3(+Berg) unc-26* recombinants showed the Bergerac pattern. One of the three *ced-3(+Berg) unc-26* recombinants showed the  
20      Bristol pattern. The genetic distance between *ced-3* and *unc-26* is 0.2 mu. Thus, *nP34* was mapped between *ced-3* and *unc-26*, about 0.1 mu on the right side of *ced-3*.

25      The flanking sequence of the 5.1 kb *EcoRI Tc1* fragment (named *nP35*) was used to probe the same set of recombinants. Two of three *ced-3(+Berg) unc-26* recombinants and two of two *ced-3 unc-26(+Berg)* recombinants showed the Bristol pattern. Thus, *nP35* was also found to be located between *ced-3* and *unc-26*,  
30      about 0.1 mu on the right side of *ced-3*.

A similar analysis using cosmid T10H5 which contains the *HindIII* RFLP (*nP36*), and cosmid B0564, which contains a *HindIII* RFLP (*nP37*), showed that *nP36* and *nP37* mapped very close or to the right of *unc-30*.

These experiments localized the *ced-3* gene to an interval of three cosmids. The positions of the RFLPs, and of *ced-3*, *unc-30* and *unc-26* on chromosome IV, and their relationships to the cosmids are shown in Figure 5 9. It was has been further demonstrated by microinjection that cosmids C37G8 and C33F2 carry the *unc-30* gene (John Sulston, personnel communication). Thus, the region containing the *ced-3* gene was limited to an interval of two cosmids. These results are 10 summarized in Figure 9.

Complementation of *ced-3* by Germline Transformation

Cosmids that were candidates for containing the *ced-3* gene were microinjected into a *ced-3* mutant to see if they rescue the mutant phenotype. The procedure 15 for microinjection was that of A. Fire (*EMBO J.* 5:2673-2680 (1986)) with modifications. *unc-31*, a mutant defective in locomotion, was used as a marker for cotransformation (Kim and Horvitz, *Genes & Dev.* 4:357-371 (1990)), because the phenotype of *ced-3* can be 20 examined only by using Nomarski optics. Cosmid C14G10 (containing *unc-31*(+)) and a candidate cosmid were coinjected into *ced-1(e1375)*; *unc-31(e928)* *ced-3(n717)* hermaphrodites, and F1 non-Unc transformants were isolated to see if the non-Unc phenotype could be 25 transmitted and established as a line of transformants. Young L1 progeny of such transformants were examined for the presence of cell deaths using Nomarski optics to see whether the *ced-3* phenotype was suppressed. Cosmid C14G10 containing *unc-31* alone does not rescue 30 *ced-3* activity when injected into a *ced-3* mutant. Table 4 summarizes the results of these transformation experiments.

As shown in Table 4, of the three cosmids injected (*C43C9*, *W07H6* and *C48D1*), only *C48D1* rescued the *ced-3*

phenotype (2/2 non-Unc transformants rescued the ced-3 phenotype). One of the transformants, nEX2, appears to be rescued by an extra-chromosomal array of injected cosmids (Way and Chalfie, Cell 54:5-16 (1988)), which 5 is maintained as an unstable duplication, since only 50% of the progeny of a non-Unc Ced(+) animal are non-Unc Ced(+). Since the non-Unc Ced(+) phenotype of the other transformant (nIS1) is transmitted to all of its progeny, it is presumably an integrated 10 transformant. L1 ced-1 animals contain an average of 23 cell corpses in the head (Table 5). L1 ced-1; ced-3 animals contain an average of 0.3 cell corpses in the head. ced-1; unc-31 ced-3; nIS1 and ced-1; unc-31 ced-3; nEX2 animals contain an average of 16.4 and 14.5 15 cell corpses in the head, respectively. From these results, it was concluded that C48D1 contains the ced-3 gene.

In order to locate ced-3 more precisely within the cosmid C48D1, this cosmid was subcloned and the 20 subclones were tested for the ability to rescue ced-3 mutants (Table 5). C48D1 DNA was digested with restriction enzymes that cut rarely within the cosmid and the remaining cosmid was self-ligated to generate a subclone. Such subclones were then injected into a 25 ced-3 mutant to look for complementation; young L1 non-Unc progeny of the transformants were examined using Nomarski optics for the presence of cell death in the head. When C48D1 was digested with BamHI and self-ligated, the remaining 14 kb subclone (named C48D1-28) 30 was found to rescue the ced-3 phenotype when injected into a ced-3 mutant (Figure 10 and Table 5). C48D1-28 was then partially digested with BglII and self-ligated. Clones of various lengths were isolated and tested for their ability to rescue ced-3.

One clone, C48D1-43, which did not contain a 1.7 kb *Bgl*III fragment of C48D1-28, was able to rescue *ced-3* (Figure 10 and Table 5). C48D1-43 was further subcloned by digesting with *Bam*HI and *Apa*I to isolate a.

5 10 kb *Bam*HI-*Apa*I fragment. This fragment was subcloned into pBSKII+ to generate pJ40. pJ40 can restore *ced-3*+ phenotype when microinjected into a *ced-3* mutant. pJ40 was subcloned by deleting a 2 kb *Bgl*III-*Apa*I fragment to generate pJ107. pJ107 was also able to rescue the

10 *ced-3* phenotype when microinjected into a *ced-3* mutant. Deletion of 0.5 kb on the left side of pJ107 could be made by *Exo*III digestion (as in pJ107del28 and pJ107del34) without affecting *ced-3* activity; in fact, one transgenic line, nEX17, restores full *ced-3*

15 activity. However, the *ced-3* rescuing ability was significantly reduced when 1 kb was deleted on the left side of pJ107 (as in pJ107del12 and pJ107del27), and the ability was completely eliminated when a 1.8 kb *Sai*II-*Bgl*III fragment was deleted on the right side of

20 pJ107 (as in pJ55 and pJ56), suggesting that this *Sai*II site is likely to be in the *ced-3* coding region. From these experiments, *ced-3* was localized to a DNA fragment of 7.5 kb. These results are summarized in Figure 10 and Table 5.

25 *ced-3* Transcript

pJ107 was used to probe a Northern blot of N2 RNA and detected a band of 2.8 kb. Although this transcript is present in 12 *ced-3* mutant animals, subsequent analysis showed that all 12 *ced-3* mutant

30 alleles contain mutations in the genomic DNA that codes for this mRNA (see below), thus establishing this RNA as a *ced-3* transcript.

The developmental expression pattern of *ced-3* was determined by hybridizing a Northern blot of RNA from

animals of different stages (eggs, L1 through L4 larvae and young adult) with the *ced-3* cDNA subclone pJ118. Such analysis revealed that the *ced-3* transcript is most abundant during embryonic development, which is 5 the period when most programmed cell deaths occur, but it was also detected during the L1 through L4 larval stages and is present in relatively high levels in young adults. This result suggests that *ced-3* is not only expressed in cells undergoing programmed cell 10 death.

Since *ced-3* and *ced-4* are both required for programmed cell death in *C. elegans*, one of the genes might act as a regulator of transcription of the other gene. To examine if *ced-4* regulates the transcription 15 of *ced-3*, RNA was prepared from eggs of *ced-4* mutants (*n1162*, *n1416*, *n1894*, and *n1920*), and a Northern blot was probed with the *ced-3* cDNA subclone pJ118. The presence of RNA in each lane was confirmed with an actin I probe. Such an experiment showed that the 20 level of *ced-3* transcript is normal in *ced-4* mutants. This indicates that *ced-4* is unlikely to be a transcriptional regulator of *ced-3*.

#### Isolation of a *ced-3* cDNA

To isolate cDNA of *ced-3*, pJ40 was used as a probe 25 to screen a cDNA library of N2 (Kim and Horvitz, Genes & Dev. 4:357-371 (1990)). Seven cDNA clones were isolated. These cDNAs can be divided into two groups: one is 3.5 kb and the other 2.5 kb. One cDNA from each group was subcloned and analyzed further. pJ85 30 contains the 3.5 kb cDNA. Experiments showed that pJ85 contains a *ced-3* cDNA fused to an unrelated cDNA; on Northern blots of N2 RNA, the pJ85 insert hybridizes to two RNA transcripts, and on Southern blots of N2 DNA, pJ85 hybridizes to more than one band than pJ40 (*ced-3*

genomic DNA) does. pJ87 contains the 2.5 kb cDNA. On Northern blots, pJ87 hybridizes to a 2.8 kb RNA and on Southern blots, it hybridizes only to bands to which pJ40 hybridizes. Thus, pJ87 contains only *ced-3* cDNA.

To show that pJ87 does contain the *ced-3* cDNA, a frameshift mutation was made in the *SacII* site of pJ40 corresponding to the *SacII* site in the pJ87 cDNA. Constructs containing the frameshift mutation failed to rescue the *ced-3* phenotype when microinjected into *ced-3* mutant animals, suggesting that *ced-3* activity has been eliminated.

#### *ced-3* Sequence

The DNA sequence of pJ87 was determined (see Figure 4; Seq. ID #18). pJ87 contains an insert of 2.5 kb which has an open reading frame of 503 amino acids (Figure 4; Seq. ID #19). The 5' end of the cDNA contains 25 bp of poly-A/T sequence, which is probably an artifact of cloning and is not present in the genomic sequence. The cDNA ends with a poly-A sequence, suggesting that it contains the complete 3' end of the transcript. 1 kb of pJ87 insert is untranslated 3' region and not all of it is essential for *ced-3* expression, since genomic constructs with deletions of 380 bp of the 3' end can still rescue *ced-3* mutants (pJ107 and its derivatives, see Figure 10).

To confirm the DNA sequence obtained from the *ced-3* cDNA and to study the structure of the *ced-3* gene, the genomic sequence of the *ced-3* gene in the plasmid pJ107 was determined (Figure 4; Seq. ID #18). Comparison of the *ced-3* genomic and cDNA sequences revealed that the *ced-3* gene has seven introns that range in size from 54 bp to 1195 bp (Figure 5A). The four largest introns, as well as sequences 5' of the

start codon, were found to contain repetitive elements. Five types of repetitive elements were found, some of which have been previously characterized in non-coding regions of other *C. elegans* genes such as *fem-1* (Spence et al., *Cell* 60:981-990 (1990)), *lin-12* (J. Yochem, personal communication), and *myoD* (Krause et al., *Cell* 63:907-919 (1990)) (Figure 4). Of these, repeat 1 was also found in *fem-1* and *myoD*, repeat 3 in *lin-12* and *fem-1*, repeat 4 in *lin-12*, and repeats 2 and 5 were novel repetitive elements.

A combination of primer extension and PCR amplification was used to determine the location and nature of the 5' end of the *ced-3* transcript. Two primers (Pex1 and Pex2) were used for the primer extension reaction. The Pex1 reaction yielded two major bands, whereas the Pex2 reaction gave one band. The Pex2 band corresponded in size to the smaller band from the Pex1 reaction, and agreed in length with a possible transcript that is trans-spliced to a *C. elegans* splice leader (Bektesh, *Genes & Dev.*, 2:1277-1283 (1988)) at a consensus splice acceptor at position 2166 of the genomic sequence (Figure 4). The nature of the larger Pex1 band is unclear.

To confirm the existence of this trans-spliced message in wild-type worms, total *C. elegans* RNA was PCR amplified using the SL1-Log5 and SL2-Log5 primer pairs, followed by a reamplification using the SL1-Oligo10 and SL2-Oligo10 primer pairs. The SL1 reaction yielded a fragment of the predicted length. The identity of this fragment was confirmed by sequencing. Thus, at least some, if not most, of the *ced-3* transcript is trans-spliced to SL1. Based on this result, the start codon of the *ced-3* message was assigned to the methionine encoded at position 2232 of the genomic sequence (Figure 4).

The DNA sequences of 12 EMS-induced *ced-3* alleles were also determined (Figure 4 and Table 3). Nine of the 12 are missense mutations. Two of the 12 are nonsense mutations, which might prematurely terminate the translation of *ced-3*. These nonsense *ced-3* mutants confirmed that the *ced-3* gene is not essential for viability. One of the 12 mutations is an alteration of a conserved splicing acceptor G, and another has a change of a 70% conserved C at the splice site, which could also generate a stop codon even if the splicing is correct. Interestingly, these EMS-induced mutations are in either the N-terminal quarter or C-terminal half of the protein. In fact, 9 of the 12 mutations occur within the region of *ced-3* that encodes the last 100 amino acids of the protein. Mutations are notably absent from the middle part of the *ced-3* gene (Figure 5).

Ced-3 Protein Contains A Region Rich in Serines

The Ced-3 protein is very hydrophilic and no significantly hydrophobic region can be found that might be a trans-membrane domain (Figure 6). The Ced-3 protein is rich in serine. From amino acid 78 to amino acid 205 of the Ced-3 protein, 34 out of 127 amino acids are serine. Serine is often the target of serine/threonine protein kinases (Edelman, *Ann. Rev. Biochem.* 56:567-613 (1987)). For example, protein kinase C can phosphorylate serines when they are flanked on their amino and carboxyl sides by basic residues (Edelman, 1987 *supra*). Four of the serines in the Ced-3 protein are flanked by arginines (Figure 4). The same serine residues might also be the target of related Ser/Thr kinases.

To identify the functionally important regions of the Ced-3 protein, genomic DNAs containing the *ced-3*

genes from two related nematode species, *C. briggsae* and *C. vulgaris* were cloned and sequenced (Figure 7; Seq. ID #20 and 21). Sequence comparison of the three *ced-3* genes showed that the non-serine-rich region of  
5 the proteins is highly conserved. In *C. briggsae* and *C. vulgaris*, many amino acids in the serine-rich region are dissimilar compared to the *C. elegans* *Ced-3* protein (Figure 7). It seems that what is important in the serine-rich region is the overall serine-rich feature  
10 rather than the exact amino acid sequence.

This hypothesis is also supported by analysis of *ced-3* mutations in *C. elegans*: none of the 12 EMS-induced mutations is in the serine-rich region, suggesting that mutations in this region might not  
15 affect the function of the *Ced-3* protein and thus, could not be isolated in the screen for *ced-3* mutants.

Table 1  
Rescue of the Ced-4  
Phenotype by Germline Transformation

Genotype	DNA Injected	Avg. No. Cell Corpses (L1 Head)	No. Animals Scored
<i>ced-1; ced-4;</i>	C10D8;	9.4	10
<i>unc-31; nEx1</i>	C14G10		
<i>ced-1; ced-4;</i>	C10D8-5	11.5	10
<i>unc-31; nEx7</i>	C14G10		
<i>ced-1; ced-4</i>	C10D8-5	11.5	10
<i>unc-31; nEx8</i>	C14G10		
<i>ced-1</i>	None	23	20
<i>ced-1; ced-4</i>	None	0.6	20

**Table 2**  
**Sites of Mutations in the *ced-4* Gene**

<u>Allele</u>	<u>Mutation</u>	<u>Nucleotide</u>	<u>Codon</u>	<u>Consequence</u>
<i>n1162</i>	C to T	1131	40	Q to ochre (TAA)
<i>n2274</i>	C to T	1428	139	R to opal (TGA)
<i>n1920</i> & <i>n2247</i>	G to A	1744	first base of 5' splice donor of intron 3	Altered splicing
<i>n2273</i>	G to A	1929	first base of 3' splice acceptor of intron 3	Altered splicing
<i>n1948</i>	T to A	2117	258	I to N
<i>n1947</i>	C to T	2128	262	Q to amber
<i>n1894</i>	G to A	3131	401	W to opal (TGA)

Nucleotide and codon positions correspond to the numbering in Figure 1.

-62-

**Table 3**  
**Sites of Mutations in the ced-3 Gene**

<u>Allele</u>	<u>Mutation</u>	<u>Nucleotide</u>	<u>Codon</u>	<u>Consequence</u>
n1040	C to T	2310	27	L to F
n718	G to A	2487	65	G to R
n2433	G to A	5757	360	G to S
n1164	C to T	5940	403	Q to termination
n717	G to A	6297	-	Splice acceptor loss
n1949	C to T	6322	412	Q to termination
n1286	G to A	6342	428	W to termination
n1129	C to T	6434	449	A to V
n1165	C to T	6434	449	A to V
n2430	C to T	6485	466	A to V
n2426	G to A	6535	483	E to K
n1163	C to T	7020	486	S to F

Nucleotide and codon positions correspond to the numbering in Figure 4.

Table 4  
Summary of Transformation Experiments  
Using Cosmids in the ced-3 Region

<u>Cosmid injected</u>	<u>No. of non-Unc transformants</u>	<u>Ced-3 phenotype</u>	<u>Strain name</u>
C43C9; C14G10	1	-	MT4302
W07H6; C14G10	3	-	MT4299
		-	MT4300
		-	MT4301
C48D1; C14G10	2	+	MT4298
		+	MT4303

Animals injected were of genotype: *ced-1(e1735); unc-31(e929)*  
*ced-3(n717)*.

Table 5

The expression of ced-3(+) transformants

Genotype	DNA injected	Average No. cell deaths in L1 head	No. Animals scored
<i>ced-1</i>	-	23	20
<i>ced-1; ced-3</i>	-	0.3	10
<i>ced-1; nIS1</i> <i>unc-31 ced-3</i>	C48D1; C14G10	16.4	20
<i>ced-1; unc-31</i> <i>ced-3; nIS1/+</i>		14.5	20
<i>ced-1; unc-31</i> <i>ced-3; nEX2</i>	C48D1; C14G10	13.2	10/14
		0	4/14
<i>ced-1; unc-31</i> <i>ced-3; nEX10</i>	C48D1-28; C14G10	12	9/10
		0	1 of 10
<i>ced-1; unc-31</i> <i>ced-3; nEX9</i>	C48D1-28; C14G10	12	10
<i>ced-1; unc-31</i> <i>ced-3; nEX11</i>	C48D1-43 C14G10	16.7	10/13
		Abnormal cell deaths	3/13
<i>ced-1; unc-31</i> <i>ced-3; nEX13</i>	pJ40; C14G10	13.75	4/4

Table 5 continued

<i>ced-1; unc-31</i>	pJ107del28,	23	12/14
<i>ced-3; nEX17</i>	pJ107del34 C14G10		
		0	2/14
<i>ced-1; unc-31</i>	pJ107del28,	12.8	9/10
<i>ced-3; nEX18</i>	pJ107del134 C14G10		
		0	1/10
<i>ced-1; unc-31</i>	pJ107del28,	10.6	5/6
<i>ced-3; nEX19</i>	pJ107del34 G14G10		
		0	1/6
<i>ced-1; unc-31</i>	pJ107del12,	7.8	12/12
<i>ced-3; nEX16</i>	pJ107del27 C14G10		

Alleles of the genes used are *ced-1(e1735)*, *unc-31(e928)*, and *ced-3(n717)*.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific 5 embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims. For example, functional equivalents of DNAs and RNAs may be nucleic acid sequences which, through the degeneracy of the genetic code, encode the 10 same proteins as those specifically claimed.

Functional equivalents of proteins may be substituted or modified amino acid sequences, wherein the substitution or modification does not change the activity or function of the protein. A "silent" amino 15 acid substitution, such that a chemically similar amino acid (e.g., an acidic amino acid with another acidic amino acid) is substituted, is an example of how a functional equivalent of a protein can be produced.

Functional equivalents of nucleic acids or proteins can 20 also be produced by deletion of nonessential sequences.

CLAIMS

1. Isolated DNA which is the ced-3 gene.
2. Isolated DNA having the nucleotide sequence of Figure 4 (Seq. ID #18).
- 5 3. Isolated DNA encoding the amino acid sequence of Figure 4 (Seq. ID #19).
4. Isolated RNA encoded by the DNA of Claim 1.
5. Isolated protein encoded by the DNA of Claim 1.
6. Isolated protein having the amino acid sequence of Figure 4 (Seq. ID #19).
- 10 7. An antibody directed against the protein of Claim 6.
8. Isolated DNA which is a mutated ced-3 or ced-4 gene having a mutation which affects the activity of the gene.
- 15 9. The DNA of Claim 8, wherein the mutated ced-4 gene is selected from the group consisting of:
  - a) n1162;
  - b) n2274;
  - 20 c) n1920;
  - d) n2247;
  - e) n2273;
  - f) n1948;
  - g) n1947; and
  - 25 h) n1894.

10. The DNA of Claim 8, wherein the mutation in ced-4 results in an alteration selected from the group consisting of:

- 5      a) Q to termination at codon 40;
- b) R to termination at codon 139;
- c) I to N at codon 258;
- d) Q to termination at codon 262;
- e) W to termination at codon 401; and
- 10     f) an alteration in mRNA splicing resulting from a change at nucleotide 6297.

11. The DNA of Claim 8, wherein the mutation in ced-4 is selected from the group consisting of:

- 15     a) C to T at nucleotide 1131;
- b) C to T at nucleotide 1428;
- c) G to A at nucleotide 1929;
- d) T to A at nucleotide 2117;
- e) C to T at nucleotide 2128; and
- f) G to A at nucleotide 3131.

20     12. The DNA of Claim 8, wherein the mutated ced-3 gene is selected from the group consisting of:

- 25     a) n1040;
- b) n718;
- c) n2433;
- d) n1164;
- e) n717;
- f) n1949;
- g) n1286;
- h) n1129;
- i) n1165;
- 30     j) n2430;
- k) n2426; and
- l) n1163.

13. The DNA of Claim 8, wherein the mutation in ced-3 results in an alteration selected from the group consisting of:

- 5       a) L to F at codon 27;
- b) G to R at codon 65;
- c) G to S at codon 360;
- d) Q to termination at codon 403;
- e) Q to termination at codon 417;
- f) W to termination at codon 428;
- 10      g) A to V at codon 449;
- h) A to V at codon 466;
- i) E to K at codon 483;
- j) S to F at codon 486; and
- 15      k) an alteration in mRNA splicing at nucleotide 6297.

14. The DNA of Claim 8, wherein the mutation in ced-3 is selected from the group consisting of:

- 20      a) C to T at nucleotide 2310;
- b) G to A at nucleotide 2487;
- c) G to A at nucleotide 5757;
- d) C to T at nucleotide 5940;
- e) G to A at nucleotide 6297;
- f) C to T at nucleotide 6322;
- g) G to A at nucleotide 6342;
- 25      h) C to T at nucleotide 6434;
- i) C to T at nucleotide 6485;
- j) G to A at nucleotide 6535; and
- k) C to T at nucleotide 7020.

15. Isolated RNA encoded by the DNA of Claim 8.

30 16. Isolated protein encoded by the DNA of Claim 8.

17. Isolated DNA which is a gene selected from the group consisting of:

- a) a gene which is structurally related to the ced-3 gene;
- 5 b) a gene which is functionally related to the ced-3 gene;
- c) a gene which is both structurally and functionally related to the ced-3 gene;
- d) a gene which is structurally related to the ced-4 gene;
- 10 e) a gene which is functionally related to the ced-4 gene; and
- f) a gene which is both structurally and functionally related to the ced-4 gene.

15 18. Isolated RNA encoded by the DNA of Claim 17.

19. Isolated protein encoded by the DNA of Claim 17.

20. An antibody directed against the protein of Claim 19.

21. A probe for identifying a gene which is structurally related to the ced-3 gene, said probe which is selected from the group consisting of:

- a) DNA having all or a portion of the nucleotide sequence of Figure 4 (Seq. ID #18);
- b) RNA encoded by the DNA of a);
- 25 c) degenerate oligonucleotides derived from a portion of the amino acid sequence of Figure 4 (Seq. ID #19); and
- d) an antibody directed against the protein of c).

22. A probe for identifying a gene which belongs to the same gene family as the *ced-3* gene, said probe which is selected from the group consisting of:

- 5        a) all or a portion of a gene which is structurally related to *ced-3*;
- b) RNA encoded by a);
- c) DNA having the consensus sequence of a conserved region between at least two other genes which belong to said gene family;
- 10      d) RNA encoded by c);
- e) degenerate oligonucleotides derived from a portion of the amino acid sequence of a protein encoded by a);
- f) degenerate oligonucleotides derived from the consensus sequence of a conserved region between the proteins encoded by at least two other genes which belong to said gene family; and
- 15      g) an antibody directed against all or a portion of a protein encoded by a).
- 20

23. A method for identifying a gene which is structurally related to a cell death gene selected from *ced-3* and *ced-4*, comprising the steps of:

- 25      a) combining DNA with a nucleic acid probe comprising said cell death gene, or a portion able to specifically hybridize to said cell death gene, under conditions suitable for specific hybridization of the nucleic acid probe to complementary sequences; and
- 30      b) detecting specific hybridization of the nucleic acid probe to the DNA, wherein specific hybridization indicates that a structurally related gene, or portion, is present in the DNA,

thereby identifying a gene which is structurally related to a cell death gene selected from ced-3 and ced-4.

24. The method of Claim 23, wherein the DNA is a gene library.

5

25. The method of Claim 23, wherein the nucleic acid probe further comprises degenerate oligonucleotides derived from the amino acid sequence of the product of the cell death gene.

10 26. A method for identifying a gene which is structurally related to a cell death gene selected from ced-3 and ced-4, comprising the steps of:

15 1) combining nucleic acid with primers comprising portions of said cell death gene under conditions suitable for polymerase chain reaction; and

2) detecting specific DNA amplification, wherein specific DNA amplification produces a structurally related gene, or portion,

20 thereby identifying a gene which is structurally related to a cell death gene selected from ced-3 and ced-4.

27. The method of Claim 26, wherein the primers further comprise degenerate oligonucleotides derived from the amino acid sequence of the product of the cell death gene.

25

28. A method for identifying a gene which is structurally related to a cell death gene selected from ced-3 and ced-4, comprising the steps of:

5           a) combining an expression gene library with an antibody directed against the protein encoded by said cell death gene under conditions suitable for specific antibody-antigen binding of the antibody to antigens expressed from the gene library; and

10          b) detecting specific antibody-antigen binding, wherein specific antibody-antigen binding indicates that a structurally related gene is present in the expression gene library, thereby identifying a gene which is structurally related to a cell death gene selected from ced-3 and ced-4.

15

29. A bioassay for identifying a cell death gene, comprising the steps of:

20          a) using a gene and a nematode selected from a nematode having reduced activity of a cell death gene and a wild-type nematode to produce a transgenic nematode; and

25          b) determining in said transgenic nematode an increase in cell deaths which occur during the development of the nontransgenic nematode, wherein an increase in cell deaths indicates the activity of a cell death gene, thereby identifying a cell death gene.

30

30. The bioassay of Claim 29, wherein the nematode underexpresses or expresses an inactivated form of a gene selected from ced-3 and ced-4.

31. The bioassay of Claim 29, wherein the gene is from an organism other than a nematode.
32. The bioassay of Claim 29, wherein the gene is a component of an expression gene library.
- 5 33. Isolated DNA which is a cell death gene identified by the bioassay of Claim 29.
34. A bioassay to identify a mutation in a cell death gene which alters the activity of the gene, comprising the steps of:
  - 10 a) using a mutated cell death gene and a nematode selected from a nematode having reduced activity of a cell death gene and a wild-type nematode to produce a transgenic nematode; and
  - 15 b) comparing cell deaths which occur during the development of the transgenic nematode having the mutated gene with those which occur in a transgenic nematode having a non-mutated gene, wherein a difference in cell deaths indicates that the mutation alters the activity of the cell death gene, thereby identifying a mutation in a cell death gene which alters the activity of the gene.
- 20 35. Isolated DNA which is a cell death gene having a mutation identified by the bioassay of Claim 34.
36. The isolated DNA of Claim 35, wherein the mutation has a result selected from the group consisting of:
  - a) inactivation of the cell death gene;

5           b) constitutive activation of the cell death gene; and

c) production of a mutated gene which does not cause cell death and which antagonizes the activity of functioning cell death genes.

10          37. A bioassay for identifying a gene which affects the activity of a cell death gene, comprising the steps of:

15           a) using a gene and a nematode containing a cell death gene to produce a transgenic nematode; and

b) determining in said transgenic nematode a difference in cell deaths from cell deaths which occur during the development of the nontransgenic nematode, wherein a difference in cell deaths indicates a gene which affects the activity of a cell death gene, thereby identifying a gene which affects the activity of a cell death gene.

20          38. The bioassay of Claim 37, wherein the cell death gene is selected from the group consisting of:

25           a) a wild-type gene;

b) an underexpressed gene;

c) a gene having reduced activity;

d) an overexpressed gene; and

e) a gene having hyperactivity.

30          39. The bioassay of Claim 37, wherein the gene is a component of an expression gene library.

40. An isolated gene identified by the bioassay of Claim 37.

41. A bioassay for identifying an agent which mimics the activity of a cell death gene, comprising the steps of:

5       a) introducing an agent into a nematode selected from a nematode having reduced activity of a cell death gene and a wild-type nematode; and

10      b) detecting an increase in cell deaths which occur in the nematode, wherein an increase indicates that the agent mimics the activity of a cell death gene,  
thereby identifying an agent which mimics the activity of a cell death gene.

42. The bioassay of Claim 41, wherein the nematode underexpresses or expresses an inactivated gene  
15      selected from ced-3 or ced-4.

43. The bioassay of Claim 42, wherein the agent is introduced into the nematode by a method selected from: microinjection, diffusion, ingestion and shooting in with a particle gun.

20     44. An agent identified by the bioassay of Claim 41.

45. A bioassay for identifying an agent which affects the activity of a cell death gene, comprising the steps of:

25      a) introducing an agent into a nematode which expresses a cell death gene; and

30      b) detecting a change in the pattern of cell deaths which occur in the development of the nematode, wherein a change indicates that the agent affects the activity of the cell death gene,

thereby identifying an agent which affects the activity of a cell death gene.

46. The bioassay of Claim 45, wherein the nematode expresses an endogenous cell death gene or a cell death gene which is a transgene.  
5
47. The bioassay of Claim 46, wherein the cell death gene is *ced-3* or *ced-4*.  
10
48. The bioassay of Claim 45, wherein the nematode overexpresses or underexpresses the cell death gene.  
10
49. The bioassay of Claim 45, wherein the nematode expresses an inactivated or constitutively activated form of the cell death gene.  
15
50. The bioassay of Claim 45, wherein the nematode underexpresses or expresses an inactivated form of a gene selected from *ced-3* and *ced-4*.  
15
51. An agent identified by the bioassay of Claim 45.  
20
52. The agent of Claim 47 which is selected from the group consisting of:
  - a) single stranded nucleic acid having all or a portion of the antisense sequence of the cell death gene which is complementary to the mRNA encoded by the gene;
  - b) DNA encoding a); and  
25
  - c) an antagonist of the cell death gene.

53. A method for altering the occurrence of cell death, comprising altering in the cell the activity of a cell death gene.

54. The method of Claim 53, wherein the cell death gene is ced-3 or ced-4.

55. The method of Claim 53, comprising exposing the cell to an agent which alters or mimics the activity of a cell death gene in the cell under conditions appropriate for activity of the agent.

10 56. The method of Claim 55, wherein the activity of the cell death gene is increased, comprising exposing the cell to an agent selected from the group consisting of:

15      a) DNA comprising the cell death gene, or active portion thereof;

      b) RNA encoded by the cell death gene, or active portion thereof;

      c) protein encoded by the cell death gene, or active portion thereof;

20      d) an agent which is structurally similar to and mimics the activity of the protein encoded by the cell death gene;

      e) DNA comprising a constitutively activated form of a cell death gene, or active portion thereof;

25      f) RNA encoded by the DNA of e), or active portion thereof;

      g) protein encoded by the DNA of e), or active portion thereof;

30      h) an agent which is structurally similar to and mimics the activity of the protein encoded by the DNA of a); and

i) an agonist of the cell death gene,  
under conditions appropriate for the activity of  
the agent.

57. The method of Claim 55, wherein the activity of  
5 the cell death gene is decreased, comprising  
exposing the cell to an agent selected from the  
group consisting of:

a) single stranded nucleic acid having all or a  
portion of the antisense sequence of the cell  
death gene which is complementary to the mRNA  
of the gene;

10 b) DNA which directs the expression of a);

c) a mutated cell death gene which does not  
cause cell death and which antagonizes the  
activity of the cell death gene;

15 d) RNA encoded by c);

e) protein encoded by c); and

f) an antagonist of the cell death gene,  
under conditions appropriate for the activity of  
20 the agent.

58. A method for reducing the proliferative capacity  
or size of a population of cells, comprising  
increasing the activity of a cell death gene in  
the cells.

25 59. The method of Claim 58, wherein the cells are  
selected from:

a) cancerous cells;

b) infected cells;

c) cells producing autoreactive  
30 antibodies; and

d) hair follicle cells.

60. The method of Claim 58, wherein the cell death gene is selected from the group consisting of:

5           a) *ced-3*;

              b) a cell death gene which is structurally related to *ced-3*; and

              c) a gene which is functionally related to *ced-3*.

61. The method of Claim 58, wherein the cell death gene is selected from the group consisting of:

10           a) *ced-4*;

              b) a cell death gene which is structurally related to *ced-4*; and

              c) a gene which is functionally related to *ced-4*.

15 62. A method for treating a condition characterized by cell deaths, comprising decreasing the activity of a cell death gene.

63. The method of Claim 62, wherein the condition is selected from the group consisting of:

20           a) myocardial infarction;

              b) stroke;

              c) degenerative disease;

              d) traumatic brain injury;

              e) hypoxia;

25           f) pathogenic infection;

              g) aging; and

              h) hair loss.

64. A method for treating a parasitic infection of a host animal, comprising administering an agent which increases the activity of a cell death gene specific to the parasite and which does not harm the host animal.

5

65. The method of Claim 64, wherein the parasite is a nematode.

66. A method for incapacitating or killing a pest, comprising increasing the activity of a cell death gene in the pest.

10

67. A method of biological containment of a recombinant organism, comprising introducing in the organism nucleic acid which is able to direct the expression of an agent which increases the activity of a cell death gene in the organism under predetermined conditions, thereby incapacitating or killing the recombinant organism.

15

68. The method of Claim 67, wherein the agent kills the recombinant organism upon completion of a desired task by the organism.

20

CLONING, SEQUENCING AND CHARACTERIZATION  
OF TWO CELL DEATH GENES AND USES THEREFOR

Abstract of the Disclosure

Described herein are genes shown to be essential

- 5 for programmed cell death in *C. elegans*, their encoded products (RNA and polypeptides), antibodies directed against the encoded polypeptides; probes for identifying structurally related genes and bioassays for identifying functionally related cell death genes
- 10 from various organisms; methods and agents for altering (increasing or decreasing) the activity of the cell death genes and, thus, of altering cell death; and uses therefor. Specifically, two genes shown to be essential for almost all of the cell deaths which occur
- 15 in the development of *C. elegans*, referred to as *ced-3* and *ced-4*, have been cloned, sequenced and characterized.

## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Declaration for Patent Application

As a named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated next to my name;

I believe I am the original, first and sole inventor (if only one name is listed) or an original, first and joint inventor (if plural names are listed in the signatory page(s) commencing at page 3 hereof) of the subject matter which is claimed and for which a patent is sought on the invention entitled

CLONING, SEQUENCING AND CHARACTERIZATION OF TWO CELL DEATH GENES  
AND USES THEREFOR

the specification of which (check one)

is attached hereto.

was filed on November 20, 1992 as  
Application Serial No. 07/979,638  
and was amended on \_\_\_\_\_ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is known by me to be material to patentability as defined in 37 C.F.R. §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s)

Priority  
Claimed

(Number)	(Country)	(Day/Month/Year filed)	<input type="checkbox"/>	<input type="checkbox"/>
----------	-----------	------------------------	--------------------------	--------------------------

Yes No

(Number)	(Country)	(Day/Month/Year filed)	<input type="checkbox"/>	<input type="checkbox"/>
----------	-----------	------------------------	--------------------------	--------------------------

Yes No

(Number)	(Country)	(Day/Month/Year filed)	<input type="checkbox"/>	<input type="checkbox"/>
----------	-----------	------------------------	--------------------------	--------------------------

Yes No

RECEIVED  
DECEMBER 11 1992  
COURT CLERK'S OFFICE

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose information known by me to be material to patentability as defined in 37 C.F.R. §1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

07/897,788 (Application Serial No.)	June 12, 1992 (Filing date)	pending (Status, patented, pending, abandoned)
<hr/>		
(Application Serial No.)	(Filing date)	(Status, patented, pending, abandoned)

As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

I also hereby grant additional Powers of Attorney to the following attorney(s) and/or agent(s) to file and prosecute an international application under the Patent Cooperation Treaty based upon the above-identified application, including a power to meet all designated office requirements for designated states.

David E. Brook	Registration No. 22,592
James M. Smith	Registration No. 28,043
Leo R. Reynolds	Registration No. 20,884
Richard A. Wise	Registration No. 18,041
Patricia Granahan	Registration No. 32,227
Mary Lou Wakimura	Registration No. 31,804
Thomas O. Hoover	Registration No. 32,470
Alice O. Carroll	Registration No. 33,542
Pei-Pei C. Smith	Registration No. 36,280

all of Hamilton, Brook, Smith and Reynolds, P.C., Two Militia Drive, Lexington, Massachusetts 02173;

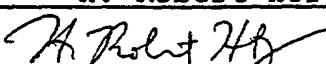
and

Send correspondence to: Patricia Granahan, Esq.  
HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
Two Militia Drive  
Lexington, Massachusetts 02173-4799

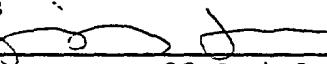
Direct telephone calls to: Patricia Granahan, Esq.  
(617) 861-6240

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

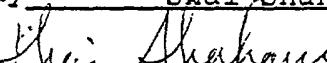
---

Full name of sole  
or first inventor H. Robert Horvitz  
Inventor's  
Signature   
Residence 80 Wendell Street, #2  
Citizenship U.S.A.  
Post Office Address same

---

Full name of second joint  
inventor, if any Junying Yuan  
Second Inventor's  
Signature   
Residence 22 Fairlee Road  
Citizenship China  
Post Office Address same

---

Full name of third joint  
inventor, if any Shai Shaham  
Third Inventor's  
Signature   
Residence 132 Commonwealth Avenue, #9  
Citizenship Israel and U.S.A.  
Post Office Address same

---

Full name of fourth joint  
inventor, if any \_\_\_\_\_  
Fourth Inventor's  
Signature \_\_\_\_\_ Date \_\_\_\_\_  
Residence \_\_\_\_\_  
Citizenship \_\_\_\_\_  
Post Office Address \_\_\_\_\_

---

PROBATE DIVISION  
MASSACHUSETTS  
SIXTY-THREE

GAATTCCGCGTCGAATCATTGTCTGGTATCGATTAGAAACCGAAACTTGTGATCGA  
 1 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 60  
 TAACAAGTCATTCAAACACGGCGAAGATGTCTATCGTATAACAGAAATTTGGAGAAAT  
 61 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 120  
 GCTCGCAAAACTCGAAATTGTCACCGATAAAATGATTAACCTGAAAGGGGCTAATGTAAGT  
 121 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 180  
 TATCTGATGTTCTACAAATTAAAAAATTGTTTTTCCAAATTAAATTTCGAAGATT  
 181 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 240  
 AACGAAAACGATTAAAATCAATAAAACGCAATAAGAGGGCTTGGCTTCTTTAAT  
 241 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 300  
 TTAAATTATAATTTCGATTGTTGATGAAGCTACAAAATGTACTGTTTGATTTG  
 301 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 360  
 AATATTGTAATTACAGGGTGGGATTCTCGCAAAATATCAGGCACAGTGGAAAGATTAGAA  
 361 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 420  
 GAAGGACGTGTGACAATCACTAAGTCAAAGAGGGAAAGGATAAAGGATTGTGATATTCA  
 421 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 480  
 CTGTTTACTCATCGCTTTAAAGAACATATGCCGATTGCCGATATTTTG  
 481 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 540  
 TTTATTAGGCCTCTCACATTCTGTACAATGTTCTACCAAAATAACTGCATTTTATCT  
 541 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 600  
 GAAAATTCAATTATTTGTCTACTTTTACTCGTTGCATTGAGATCAGCATATCTT  
 601 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 660  
 CCGGTCTATTTATATTCAACGATTTTATAAATTAGTACTCCTCATGTTAATTTCATT  
 661 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 720  
 TTATCTGTAAGCTTACTGTATTTTAAATCTTCTTCTATCTGATTATACAA  
 721 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 780  
 TGTTCTTACTCATTCAGGTATTTTATGCCCTACAATTATGCACATTGGGCTT  
 781 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 840  
 GGAGATTTATCCTCTATATTACATGCCGTGTTTAAAGGATATAATGTTAACAAATA  
 841 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 900  
 ATTTTTATCAATGCTATTGTATATTCTCCAGCTAACGGTGTTCGAAACATCACCTA  
 901 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 960  
 GCATTTAAAATTCAACAAATCTGCTTCCCTATAATCAAGAAGATTTCAGATGCTCT  
 961 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1020  
 M L C  
 !  
 GCGAAATCGAATGCCGGCTTGAGCACGGCACACACGAGGCTCATCCACGACTTGAAC  
 1021 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1080  
 E I E C R A L S T A H T R L I H D F E P  
 ,  
 10 20  
 ↑ n1162  
 CACGTGACGCATTGACTTATTTAGAAGGCAAAACATTTACAGAAAGATCATTCTGAAC  
 1081 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1140  
 R D A L T Y L E G K N I F T E D H S E L  
 ,  
 30 40  
 TTATCAGTAAAATGTCACACTGCCCTCGAGAGGATGCCAATTTCTTCAATCTATCGAC  
 1141 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1200  
 I S K M S T R L E R I A N F L R I Y R R  
 ,  
 50 60  
 GTCAAGCTCTGAACCTGGACCACACTCATCGACTTTCAACTACAACATCAAAGTCACC  
 1201 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1260  
 Q A S E L G P L I D F F N Y N N Q S E L  
 ,  
 70 80  
 TTGCTGATTCTCGAAGACTACATCGATTTGCGATAATGAGCCAGATCTACTTCGTC  
 1261 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1320  
 A D F L E D Y I D F A I N E P D L L R P  
 ,  
 90 100  
 CACTAGTGTATTGCTCCACAATTCCCACAAATGCTCGATAGGAAACTATTGCTTGGGA  
 1321 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1380  
 V V I A P Q F S R Q M L D R K L L G N  
 , , , , , , , , , , , , , , , , , , ,  
 110 120

FIGURE 1

T n2274

ATGTTCCAAAACAAATGACATGCTATTCGAGAGTATCACGTGGATCGACTGATCAAAA  
 1381 V P K Q M T C Y I R E Y H V D R V I K K  
 130 140

Intron 1

AGCTCGACGAGATGTGATTAGGTGAGAAAATGGAAGCTCTCGTGTATTATAATC  
 1441 L D E M C D L D  
 150

TTGCTTAAACTTCAGACTCCTTTCTGTTCTACACGGCCGAGCTGGATCCGGAAAAT  
 1501 S F F L F L H G R A G S G . K S  
 160

Intron 2

CAGTAATTGCATACAAGCTTTCGAAATCTGACCAACTTATTGAAATGTGAGTGGTAT  
 1561 V I A S Q A L S K S D Q L I G I  
 170 180

TATCTGAATCTACGGATCTTCATTCTATTACAGAAATTATGATTCAATCGTTGGCTCAA  
 1621 N Y D S I V W L K  
 190

AGATAGTGGAACAGCTCCAAAATCTACATTGATTATTACGGATATTCGCTGATGCT  
 1681 D S G T A P K S T F D L F T D I L L M L  
 200 210

A n1920/n2247

Intron 3

AAATGACTGAATAGAGTGCATGAAACATTCAAGCATGATTTGAAATTATGAAAATTGAA  
 1741 K  
 CCTGGTTAGCTTTAATTGATATTCTGACGCCATGGCATGTTGTGTGTTGAAGACG  
 1801 AGCCCCTGTTGTGAGCGACACGGATGACTCCGATCCGACTTCATTAACCGTGT  
 1861

A n2273

TCTTTCAAGAAGCGAAGACGATCTCTCAATTCCCATGGGAGCATGTCACGTCACT  
 1921 S E D D L L N F P S V E H V T S V  
 220

Intron 4

TGTACTCAAAAGGATGTAAGTTGCTGCCGATTCTGGTACAATATCTTAAATTATTGGT  
 1981 V L K R M  
 230

TTTAGATCTGCAACGCACTCATGATCGTCCAAATACTTATTCTGATGACGTA  
 2041 I C N A L I D R P N T L F V F D D V  
 240 250

A n1948 T n1947

GTTCAAGAAGAAACAATTGGCTCAGGAGCTACGTCTCGATGCTTGTAACACT  
 2101 V Q E E T I R W A Q E L R L R C L V T T  
 260 270

CGTGACCTGGAAATATCAAATGCTGCTCTCAAACATGCGAATTGAGTGCACATCA  
 2161 R D V E I S N A A S Q T C E F I E V T S  
 280 290

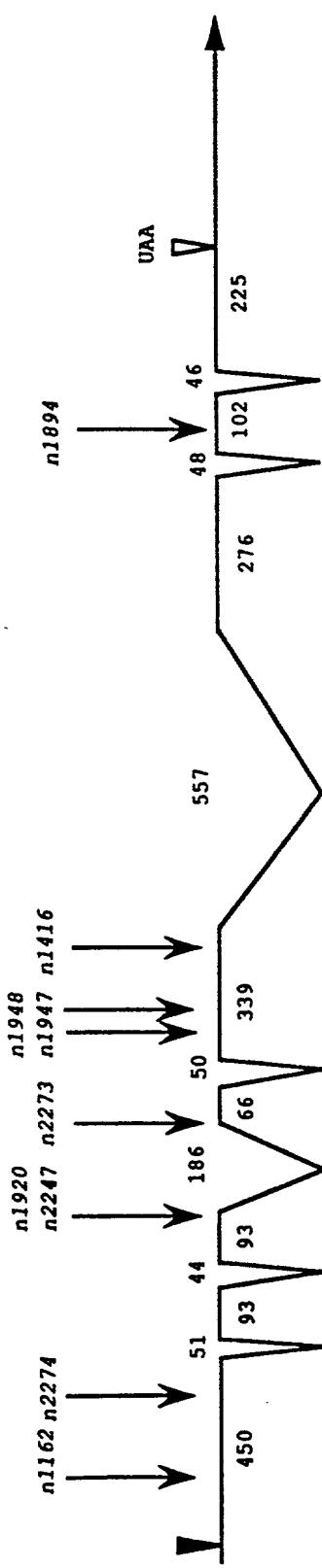
FIGURE 1

FIGURE 1

TTTGACATTCAAAATTGATCATATTATCCATAIGTTCTTGAACACGTCGTGATGCAC  
 3301 L T F K I D H I I H M F L K H V V D A Q  
 450 460  
 | Intron 7 |  
 AAACATCGCGTATGCTGAAAATGCTCAACTTCAATTAAATTAAATTTCAGAAT  
 3361 T I A N  
 CGAACATCTCAATCTCGACCACCGTCTCTTGAATAGGAAACAATAATGTACAGTACCG  
 3421 G I S I L E Q R L L E I G N N N V S V P  
 470 480  
 GAGCGACATATACCATCACATTCCAAAAATTCCGTCGTCATCAGCCAGTGAGATGTAT  
 3481 E R H I P S H F Q K F R R S S A S E M Y  
 500 510  
 CCAAAAACATACAGAAGAAAATGTGATCCGTCCTGAAGACTTCCCAAAGTCATGCAATTG  
 3541 P K T T E E T V I R P E D F P K F M Q L  
 520 530  
 CACCAGAAAATCTATGACTCCCTCAAAATTGGCATGCTTTAAACCTATCGTGTACA  
 3601 H Q K F Y D S L K N F A C C \*  
 540  
 ATATTGCCGTATATTCCCTCGAAATACGTTTACTTTCCGACGAGTTTCTCATT  
 3661 TTTTCATTTGACTTGTTTATTTCTCCAAAATTTCAGATCTATCCCAAATGTTCTTA  
 3721 AATTAATGTTTCTACAGATACTCAACACATCTGTTCATCTCATCCTGTTTTTT  
 3781 TTTCAAATATATTCAAGTTCTTTATAATTAAATTAATCGAATTAAACATTACAGTAA  
 3841 AGAATTCTGGACTATTATTTATCCATCCAAATGATTATCCCTATTGTCGAAAC  
 3901 TTCCAAATTGATCATTTAAACACGCCCTATTAAATTGAAAGTCGTACTTTAGTCG  
 3961 AACATGAAGTAAGTTATTCTGTGTTCTAAATTCAAAGTGCATTCCAAAAGGACATTG  
 4021 ATGAGTTTCAAGAAAACGTAATTTCACATTCTTCAAGTTGATGTTGAGATGTTGAT  
 4081 TTCTTCTCTGTTGGCGTCATTACTAATTGCTTCTGCTTCACATTATCGAGATT  
 4141 TTGCCATCAATGGAGTCCATCTAGACCGATAGCAGTCTCATATCATTATCCCTGTATA  
 4201 TTGTACTGTTCACTATTAACTTATCGATTACGTACTATATTCACTGGTTCACTGTT  
 4261 TCGGTCAATGGGTGACACGTGCTCGACGANNAATTTCACCGAACCGAATCTCCTAGTCA  
 4321 CTTATCAACCAAGAGCCCTCACCCATG  
 4381 4407

FIGURE 1

FIGURE 2



	10	12	14	18	21								
	X	Y	Z	-X	-Z								
Calcium-binding loop consensus	D	N	S	T	E								
	D	N		S									
		D		E									
				Q									
				D									
				N									
EF-hand consensus	O	*	O	G	*	*	O	*	*	E			
ced-4 sequence 1	Y	N	N	Q	S	H	L	A	D	F	L	E	
sequence 2	S	L	E	I	D	E	C	Y	D	F	L	E	
Parvalbumin (carp) (hake) (ray)	D	Q	D	K	S	G	F	I	E	E	D	E	
SCBP ( <i>Amphioxus I</i> )	D	Q	D	K	D	D	F	I	G	E	D	E	
ICaBP (bovine)	D	S	D	G	D	H	K	I	G	V	D	E	
Troponin C (rabbit)	D	A	D	G	G	G	D	I	S	V	K	E	
	D	E	D	G	S	G	T	I	D	F	E	E	
	D	R	N	A	D	G	Y	I	D	A	E	E	
	D	K	N	N	D	G	R	I	D	F	D	E	
Calmodulin (bovine)	D	K	D	G	N	G	T	I	T	T	K	E	
Trypsinogen	L	G	E	D	N	I	N	V	V	E	G	N	E
Fibrinogen	D	N	D	N	D	K	F	E	G	N	C	A	E
Villin	G	V	D	P	S	R	K	E	N	H	L	S	
GBP	D	L	N	K	D	G	Q	I	Q	_	I	E	

FIGURE 3

## ced-3 Genomic Sequence

1	AGATCTGAAATAAGGTGATAAATTAAATTAAAGTGTATTCTGAGGAATTGACTGT	
61	TTTAGCACATTAAATCTTGTTCAGAAAAAGTCCAGTTCTAGATTTCCGTCTTA	60
121	TTGTCGAATTAAATATCCCTATTATCACTTTCATGCTCATCCTCGAGCGGCACGTCCTC	120
181	AAAGAATTGTGAGAGCAAACGCCCTCCATTGACCTCACACTCAGCCGCCAAAACAAAC	180
241	GTTCGAACATTCGTGTGTTGTGCTCCTTCCGTTATCTGCAGTCATCTTGTGCGTT	240
301	TTTTCTTGTTCTTTGTTGAACGTGTTGCTAACATTATTACATCAATTGAAGAAAA	300
361	GGCTCGCCGATTTATTGTTGCCAGAAAGATTCTGAGATTCTGAAGTCGATTTATAATA	360
421	TTAACCTTGGTTTTGCATTGTTCGTTAAAAAAACACTGTTATGTGAAAAACGAT	420
481	TAGTTTACTAATAAAACTACTTTAACCTTACCTTACCTCACCGCTCCGTGTTCATG	480
541	GCTCATAGATTTGATACTCAAATCCAAAAATAATTACGAGGGCAATTATGTGAAA	540
601	CAAAACAATCTAAGATTCCACATGTTGACCTCTCCGGCACCTTCTCCTAGCCCC	600
661	ACCACTCCATCACCTCTTGGCGGTGTTCTCGAAACCCACTTAGGAAAGCAGTGTGTAT	660
721	CTCATTGGTATGCTCTTTCGATTAGCTCTTGTGCAATTCAATGCTTTAAC	720
781	AATCCAATCGCATTATATTGTCATGGAGGCAAATGACGGGGTGGAAATCTTAGATGA	760
841	GATCAGGAGCTTCAGGTAAACGCCCGGTTCATTTGTACACATTTCATCATTTCCT	840
901	GTCGTCCTGGTATCCTCAACTGTCCCAGTTGTTTCGGTACACTCTTCCGTGATGC	900
961	CACCTGTCTCCGTCTCAATTATCGTTAGAAATGTGAACGTGTCAGATGGGTGACTCATA	960
1021	TTGCTGCTGCTACAAATCCACTTTCTCATCGGAGCTTACGGAGGCCATCATAAAC	1020
1081	TTTTTTTCCGCGAAATTGCAATAAACCGGCCAAAACCTTCTCAAATTGTTACGCAA	1080
1141	TATATACAATCCATAAGAATATCTCTCAATGTTATGTTCTTCGCGACTTTCTCT	1140
1201	TCGTGTGCTAACATCTTATTTTATAATTTCGCTAAATTCCGATTGGTACTTGTGAGTATTA	1200
1261	ATTTATCGAAAATTATCATAATAGCACCGAAAACACTACTAAAATGGTAAAAGCTCCTT	1260
		1320

### Repeat 1

1321	----- TAAATCGGCTCGACATTATCGTATTAAGGAATCACAAATTCTGAGAATGCGTACTGCGC	1380
------	--	------

1381	----- AACATATTGACGGCAAATATCTCGTAGCGAAAACACTACAGTAATTCTTAAATGACTAC	1440
------	--	------

### Repeat 1

1441	-----> TGTAGCGCTTGTGTCGATTACGGGCTCAATTGAAAATAATTTCGAAATT	1500
------	---	------

FIGURE 4

1501 TGATAACCGTAAATCGTCACAACGCTACAGTAGTCATTAAAGGATTACTGTAGTTCTA  
 1561 GCTACGAGATATTTGCGGCCAATATGACTGTAATACGCATTCTCTGAATTTGTGTT  
 1621 TCCGTAATAATTCACAAGATTTGGCATTCACTTTAAAGGCGCACAGGATTATTCCA  
 1681 ATGGGTCTCGGCACGCAAAAGTTGATAGACTTTAAATTCTCCTTGCACTTTAATTC  
 1681 AATTACTAAAATTTCGTGAATTTCTGTTAAAATTTAAAATCAGTTTCTAATATT  
 1741 TTCCAGGCTGACAAACAGAAACAAAAACACAAACACATTAAAATCAGTTTCAAAT  
 1801 TAAAAAATAACGATTCTCATTGAAAATTGTGTTTATGTTTGCAGAAATAAGAGAACT  
 1861 GATTCAAAACAATTTAACAAAAAAACCCCCAAATTGCCAGAAATCAAGATAAAAAAA  
 1921 TTCAAGAGGGTCAAAATTTCCGATTTACTGACTTTCACCTTTTTCGTAGTTCAAGT  
 1981 GCAGTTGGAGTTTGACGAAAATCGATAAAAATTACTCAAATCG  
 2041 AGCTGAATTGAGGACAATGTTAAAAAAACACTATTTCATAATTCACTCAT  
 2101  
 2161  
 2221  
 2281  
 2341  
 2401  
 2461  
 2521

FIGURE 4

TTTTAAAGTCGGCGAAAGCAAGGTCTACGGAAAAAGAGGCCGATCGTAATT  
 2581 -----+-----+-----+-----+-----+-----+-----+-----+ 2640  
 GCAACCACCGCACGGTTTTCTCCGAAAATCGAAATTATGCACCTTCCAAATAT  
 2641 -----+-----+-----+-----+-----+-----+-----+-----+ 2700  
 TTGAAGTAAATATTTTACTGAAAGCTCGAGTATTATTTTTAACACTA  
 2701 -----+-----+-----+-----+-----+-----+-----+-----+ 2760  
 ATTTTCGTGGCGCAAAGGCCATTGTAGATTGCCGAAAATCTTGTACACACAC  
 2761 -----+-----+-----+-----+-----+-----+-----+-----+ 2820  
  
 ACACACATCTCCTTCAAATATCCCTTTCCAGTGTGACTCGAATGCTGTCGAATTCGA  
 2821 -----+-----+-----+-----+-----+-----+-----+-----+ 2880  
 V D S N A V E F E  
 100  
  
 GTGTCCAATGTCACCGCAAGCCATCGTCGGAGCCGCGCATGGAGCCCCGGCTACAC  
 2881 -----+-----+-----+-----+-----+-----+-----+-----+ 2940  
 C P M S P A S H R R S R A L S P A G Y T  
 110 120  
  
 TTCACCGACCCGAGTTCACCGTGACAGCGTCTTCAGTGTCACTCATTCACCTTATCA  
 2941 -----+-----+-----+-----+-----+-----+-----+-----+ 3000  
 S P T R V H R D S V S S V S S F T S Y Q  
 130 140  
  
 GGATATCTACTCAAGAGCAAGATCTCGTCTCGATCGCGTGCACTTCAATTGATCG  
 3001 -----+-----+-----+-----+-----+-----+-----+-----+ 3060  
 D I Y S R A R S R S R A L H S S D R  
 150 160  
  
 | intron 3  
 ACACAATTATTCACTCCCTCCAGTCAACGCATTCCCAGCCAACCTTGTATGTTGATGCG  
 3061 -----+-----+-----+-----+-----+-----+-----+-----+ 3120  
 H N Y S S P P V N A F P S Q P S  
 170  
  
 Repeat 1  
  
 AACACTAAATTCTGAGAATGCGCATTACTCACACATATTGACGCGCAAATATCTCGTAGC  
 3121 -----+-----+-----+-----+-----+-----+-----+-----+ 3180  
  
 GAAAAATACAGTAACCTTAAATGACTATTGAGTGTGCGATTACGGGCTCGATTTCG  
 3181 -----+-----+-----+-----+-----+-----+-----+-----+ 3240  
  
 -->  
 AAACGAATATATGCTCGAATTGTGACAACGAATTAAATTGTCATTTGTGTTTCTT  
 3241 -----+-----+-----+-----+-----+-----+-----+-----+ 3300  
  
 Repeat 1  
 <--  
 TTGATATTGGATCAATTAATAATTATTCCTCGAACAGACACCAGCGCTACAGTACT  
 3301 -----+-----+-----+-----+-----+-----+-----+-----+ 3360  
  
 CTTTTAAAGAGTTACAGTAGTTCTCGCTCAAGATATTGAAAAGAATTAAACATT  
 3361 -----+-----+-----+-----+-----+-----+-----+-----+ 3420  
  
 TGAAAAAAATCATCTAACATGTGCCAACGCTTTCAAGTTCGCAGATTGGAA  
 3421 -----+-----+-----+-----+-----+-----+-----+-----+ 3480

### FIGURE 4

Repeat 2

3481	-----+-----+-----+-----+-----+-----+	3540
3541	-----+-----+-----+-----+-----+-----+	3600
3601	-----+-----+-----+-----+-----+-----+	3660
3661	-----+-----+-----+-----+-----+-----+	3720
3721	-----+-----+-----+-----+-----+-----+>	3780
3781	<-----+-----+-----+-----+-----+-----+	3840
3841	-----+-----+-----+-----+-----+-----+	3900
3901	-----+-----+-----+-----+-----+-----+	3960
3961	-----+-----+-----+-----+-----+-----+	4020
4021	-----+-----+-----+-----+-----+-----+	4080
4081	-----+-----+-----+-----+-----+-----+	4140
4141	-----+-----+-----+-----+-----+-----+	4200
4201	-----+-----+-----+-----+-----+-----+	4260
4261	-----+-----+-----+-----+-----+-----+	4320

S A N S S F  
180

FIGURE 4

TCACCGGATGCTCTCTCGGATAACAGTTCAAGTCGTAAATCGCTCATTCAAGCAAAGCTT  
 4321 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 4380  
 T G C S S L G Y S S S R N R S F S K A S  
 190 200

CTGGACCAACTCAATACATATTCCATGAAGAGGATATGAACCTTGTCGATGCACCAACCA  
 4381 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 4440  
 G P T Q Y I F H E E D M N F V D A P T I  
 210 220

TAAGCCGTGTTTCGACGAGAAAACCATGTACAGAACTTCTCGAGTCCTCGTGGATGT  
 4441 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 4500  
 S R V F D E K T M Y R N F S S P R G M C  
 230 240

GCCTCATATAAATAATGAACACTTGAGCAGATGCCAACACGGAATGGTACCAAGGCCG  
 4501 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 4560  
 L I I N N E H F E Q M P T R N G T K A D  
 250 260

ACAAGGACAATCTTACCAATTGTTCAAGATGCATGGCTATACGGTTATTGCAAGGACA  
 4561 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 4620  
 K D N L T N L F R C M G Y T V I C K D N  
 270 280

| intron 4

ATCTGACGGGAAGGGTACGGCGAAATTATATTACCCAAACCGGAAATTGCCCATTGCG  
 4621 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 4680  
 L T G R

Repeat 3

CCGAAAATGTGGCGCCGGTCTCGACACGACAATTGTGTTAAATGCAAAATGTATAAT  
 4681 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 4740  
 TTTGCAAAAACAAAATTGAACTTCCCGCGAAAATGATTACCTAGTTGCAATTTC  
 4741 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 4800  
 GTTTTTCCGGCTACATTATGTGTTTTCTTAGTTTCTATAATATTGATGTA  
 4801 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 4860  
 ACCGTTTGAAATTTCAGACAATTTCGCATACAAAACCTTGATAGCACGAAATCAATT  
 4861 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 4920  
 TTCTGAATTTCAAAATTATCCAAAAATTGACAATTAAAATTGTGAAATTGGCAAAC  
 4921 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 4980  
 GGTGTTCAATATGAAATGTATTTAAAAACTTAAACCACTCCGGAAAGCAATAA  
 4981 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 5040  
 AAATCAAAACACGTACAATTCAAATTCAAAGTTATTCATCCGATTGTTATTG  
 5041 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 5100  
 CAAAATTGAAAAAAATCATGAAGGATTAGAAAAGTTTATAACATTTCAGATT  
 5101 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 5160  
 TCAAAATTTTTAAACAAATCGAGAAAAGAGAATGAAAATCGATTAAAATATCC  
 5161 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 5220

Repeat 3

<-----  
 ACAGCTTCGAGAGTTGAAATTACAGTACTCCTTAAAGGCACACCCATTGCATTGG  
 5221 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 5280

ACCAAAAATTGTCGTGTCGAGACCAGGTACCGTAGTTTGTGCAAAATTGCA  
 5281 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 5340  
 TGGACAATAACCTCTTAAATCACAAAAAGTAAAATTGAAATCTTCGAAAAGCCA  
 5341 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 5400

FIGURE 4

ATTCAAAAAAAAGTCGAATTCGATTTTTGGTTTTGGTCCCACCAAA  
 5401 -----+-----+-----+-----+-----+-----+-----+ 5460  
 AAATCAATTCTGCACAAACCCGAAAAATTCCAGCCTGTCCT  
 5461 -----+-----+-----+-----+-----+-----+-----+ 5520  
 |  
 AATGTAACTGATATTAATTCCAGGGAATGCTCCTGACAATTGAGACTTGCCAAAC  
 5521 -----+-----+-----+-----+-----+-----+-----+ 5580  
 G M L L T I R D F A K H  
 290 300  
 ACGAATCACACGGAGATTCTGCATACTCGTGATTCTACACGGAGAGAGAATGTGA  
 5581 -----+-----+-----+-----+-----+-----+-----+ 5640  
 E S H G D S A I L V I L S H G E E N V I  
 310 320  
 TTATTGGAGTTGATGATACCGATTAGTACACAGAGATATGATCTTCTAACCGCG  
 5641 -----+-----+-----+-----+-----+-----+-----+ 5700  
 I G V D D I P I S T H E I Y D L L N A A  
 330 340  
 A(n2433)  
 CAAATGCTCCCCGTCTGGCGAATAAGCCGAAATCGTTTGTGCAGGCTTGTGAGGCG  
 5701 -----+-----+-----+-----+-----+-----+-----+ 5760  
 | | intron 5  
 N A P R L A N K P K I V F V Q A C R G E  
 350 360  
 GTTCGTTTTATTTAATTTAATATAAATTTAAATAATTCAATTTCAGAACGTC  
 5761 -----+-----+-----+-----+-----+-----+-----+ 5820  
 R R  
 GTGACAATGGATTCCAGTCTGGATTCTGCGACGGAGTCCTGCATTCTCGTCGTG  
 5821 -----+-----+-----+-----+-----+-----+-----+ 5880  
 D N G F P V L D S V D G V P A F L R R G  
 370 380  
 T(n1165)  
 GATGGGACAATCGAGACGGGCCATTGTTCAATTCTTGGATGTGCGGGCGCAAGTTC  
 5881 -----+-----+-----+-----+-----+-----+-----+ 5940  
 W D N R D G P L F N F L G C V R P Q V Q  
 390 400  
 | intron 6  
 AGGTTGCAATTAAATTCTTGAATGAGAATATTCCCTCAAAAATCTAAAATAGATT  
 5941 -----+-----+-----+-----+-----+-----+-----+ 6000  
 ATTCCAGAAAGTCCCGATCGAAAAATTGCGATATAATTACGAAATTGTGATAAAATGAC  
 6001 -----+-----+-----+-----+-----+-----+-----+ 6060  
 Repeat 4  
 AAACCAATCAGCATCGTCGATCTCCGCCACTTCATCGGATTGGTTGAAAGTGGCGGA  
 6061 -----+-----+-----+-----+-----+-----+-----+ 6120  
 >  
 GTGAATTGCTGATTGGTCGCAGTTTCAGTTAGAGGAAATTAAAAATCGCCTTTCGA  
 6121 -----+-----+-----+-----+-----+-----+-----+ 6180  
 AAATTAAAAATTGATTTTCATAAATTTTCGAAAAATTCCGATTATTTATATTCTT  
 6181 -----+-----+-----+-----+-----+-----+-----+ 6240

FIGURE 4

A(n717)

6241 GGAGCGAAAGCCCCGTCTGAAACATTTAAATGATAATTAAATAAATTTGCAGCAA  
-----+-----+-----+-----+-----+-----+-----+-----+  
Q

T(n1949)

6301 GTGTGGAGAAAGAACGCCGAGCAAGCTGACATTCTGATTGATCGAACGACAGCTCAA  
-----+-----+-----+-----+-----+-----+-----+  
V W R K K P S Q A D I L I R Y A T T A Q  
410 420

A(n1286)

6361 TATGTTCTGGAGAAACAGTGCTCGTGGATCATGGTCATTCAAGCCGTCTGTGAAGTG  
-----+-----+-----+-----+-----+-----+  
Y V S W R N S A R G S W F I Q A V C E V  
430 440

T(n1129, n1164)

6421 TTCTCGACACACGCAAAGGATATGGATGTTGAGCTGACTGAAGTCATAAGAAG  
-----+-----+-----+-----+-----+-----+  
F S T H A K D M D V V E L L T E V N K K  
450 460

T(n2430) A(n2426)  
| | intron 7

6481 GTCGCTTGTGGATTTCAGACATCACAGGGATCGAATATTTGAAACAGATGCCAGAGGTA  
-----+-----+-----+-----+-----+-----+  
V A C G F Q T S Q G S N I L K Q M P E  
470 480

Repeat 5

6541 CTTGAAACAAACAATGCATGTCTAACCTTAAAGGACACAGAAAAATAGGCAGAGGCTCCT  
-----+-----+-----+-----+-----+-----+  
6600

----->

6601 TTTGCAAGCCTGCCGCGTCACCTAGAATTTAGTTTAGCTAAATGATTGATTT  
-----+-----+-----+-----+-----+-----+  
6660

6661 GAATATTTATGCTAATTTTGTGCTTAAATTTGAAATAGTCACTATTTATCGGGTT  
-----+-----+-----+-----+-----+-----+  
6720

6721 CCAGTAAAAATGTTATTAGCATTGGATTTACTGAAAACGAAAATTGTTAGTTTC  
-----+-----+-----+-----+-----+-----+  
6780

6781 AACGAAATTATCGATTTAAATGTAACGAAAATAGCGAAAATTACATCAACCATCAA  
-----+-----+-----+-----+-----+-----+  
6840

6841 GCATTTAAGCCAAATTGTTAACCTATTAAATTCAAAGTTGTCCACGAGTATT  
-----+-----+-----+-----+-----+  
6900

Repeat 5

6901 <----- ACACGGTTGGCGCGCGCAAGTTGCAAAACGACGCTCCGCCTTTCTGTGCGGCTT  
-----+-----+-----+-----+-----+-----+  
6960

T(n1163)

6961 -----  
GAAAACAAGGGATCGGTTAGATTTCCCAAAATTAAATTCAAGATGACATC  
-----+-----+-----+-----+-----+-----+  
M T S

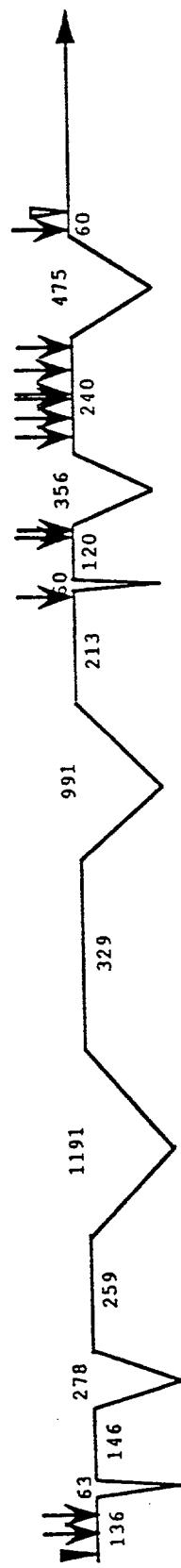
7020

FIGURE 4

7021	CCGCCTGCTCAAAAAGTTCTACTTTGGCCGGAAGCACGAAACTCTGCCGTCTAAAATTC -----+-----+-----+-----+-----+-----+ R L L K K F Y F W P E A R N S A V * 490 500	7080
7081	ACTCGTGATTCAATTGCCAATTGATAATTGCTGTATCTTCTCCCCAGTTCTCTTCGC -----+-----+-----+-----+-----+ CCAATTAGTTAAAACCATGTGATATTGTTATCCTATACTCATTCACTTATCATTCT	7140
7141	ATCATTTCTCTTCCCATTTCACACATTCCATTCTCTACGATAATCTAAAATTATGAC	7200
7201	GTTTGTGTCTCGAACGCATAATAATTAACTCGTTTGAAATTGATTAGTTGTTGT	7260
7261	GCCCAGTATATGTTATGTAATGCTTCTATCAACAAAATAGTTCATAGATCATCACC	7320
7321	CCAACCCCACCAACCTACCGTACCATATTCAATTGGCCGGAAATCAATTTCGATTAATT	7380
7381	TTAACCTATTTTCGCCACAAAAATCTAATATTGAATTAAACGAATAGCATTCCCATC	7440
7441	TCTCCCGTGCCGGAATGCCTCCGGCCTTTAAAGTTGGAACATTGGCAATTATGTAT	7500
7501	AAATTTGTAGGTCCCCCCCACATTCCGCCATCATCTCAAATTGCATTCTTTTCG	7560
7561	CCGTGATATCCGATTCTGGTCAGCAAAGATCT	7620
7621	-----+-----+-----+-----+-----+-----+ 7653	

FIGURE 4

FIGURE 5A



## *ced-3* Mutations are Clustered

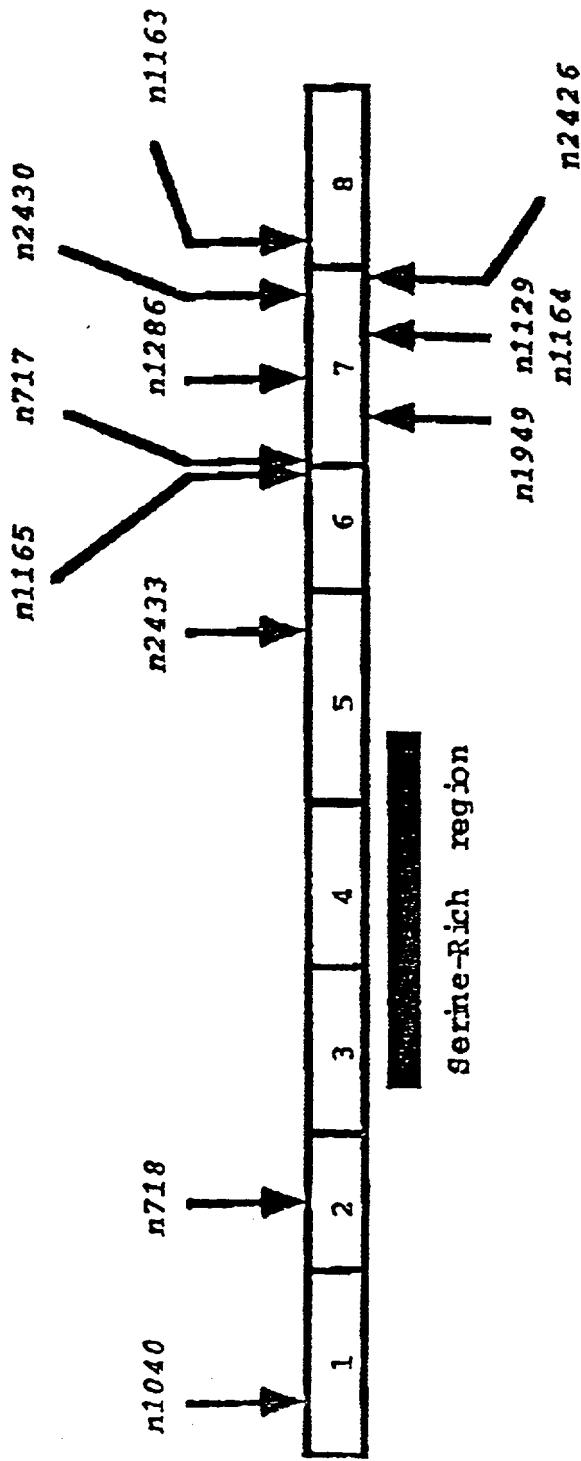


FIGURE 5B

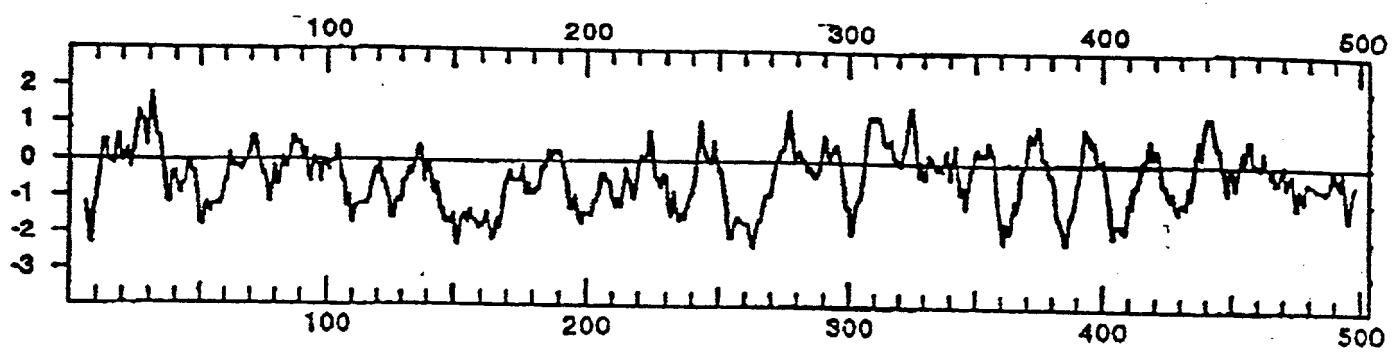


FIGURE 6

Lines

1 01 MMRQDRRSLLERNIMMFSSHLKVDEILEVLIAKQVLNSDNGDMINSCGTV 50  
2 .....W.....LE...K.QA.L..D.....V....R.E  
3 .....T.V.S.I..R.....M....

1 51 REKRREIVKAVQRPGDVAFDAFYDALRSTGHEGLAEVLEPLARSVDNAV 100  
2 ..DNEK.....R..E.....D...ND..D..M..S.P ..P.  
3

1 101 EFECPMSPASHRRSRALSPAGYTSPTRVHRDSVSSVSSFTS\_YQDIYSRA 149  
2 PM.....S.....P .A.....I.....T...V....  
3 .....S

1 150 RSRSR\_SRALHSSDRHNYSPPPVNAPSQFSSANSSFTGCSSLGYSSSRN 198  
2 ..S..S..P.Q.....M.AA\_TS.....A.....  
3 T...\_\_P.T.....V..S\_.S.Q...A.....S.....T

1 199 RSFSKASGPTQYIFHEEDMNFDAPTISRVFDEKTMYRNFSSPRGMCLI 247  
2 .....T.AQS.....Y.....H.....L...  
3 ..Y....AHS.....Y.....H.....T...L...

1 248 INNEHFEQMPTRNGTKADKDNLTNLFRCMGTICKDNLTGRGMLLTIRD 297  
2 .....I.....E..S...S  
3 .....P....IS.....I.H.....M....

1 298 FAKHESHGDSAILVILSHGEENVIIIGVDDIPISTHEIYDLLNAANAPRLA 347  
2 ..GRNDM.....VSVNV.....  
3 ...N.T.....VSVNV...X...

1 348 NKPKIVFVQACRGERRDNGFPVLDSDGVPAFLRRGWDNRDGPLFNLG 397  
2 .....L.....SLI.....  
3 .....L.....V.....LI.....KG...

1 398 VRPQVQQVWRKKPSQADILIRYATTAQYVSWRNSARGSWFIQAVCEVFST 447  
2 .....M..A.....L  
3 .....A.....A.....L

1 448 HAKDMDVVELLTEVNKKVACGFOTSQGSNILKQMPEMTSRLKKFYFWPE 497  
2 .....L.....  
3 .....A.....L.....

1 498 ARN SAV 503  
2 DRG..  
3 .....D..RS...

Line 1 *C. elegans*  
Line 2 *C. briggsae*  
Line 3 *C. vulgaris*

FIGURE 7

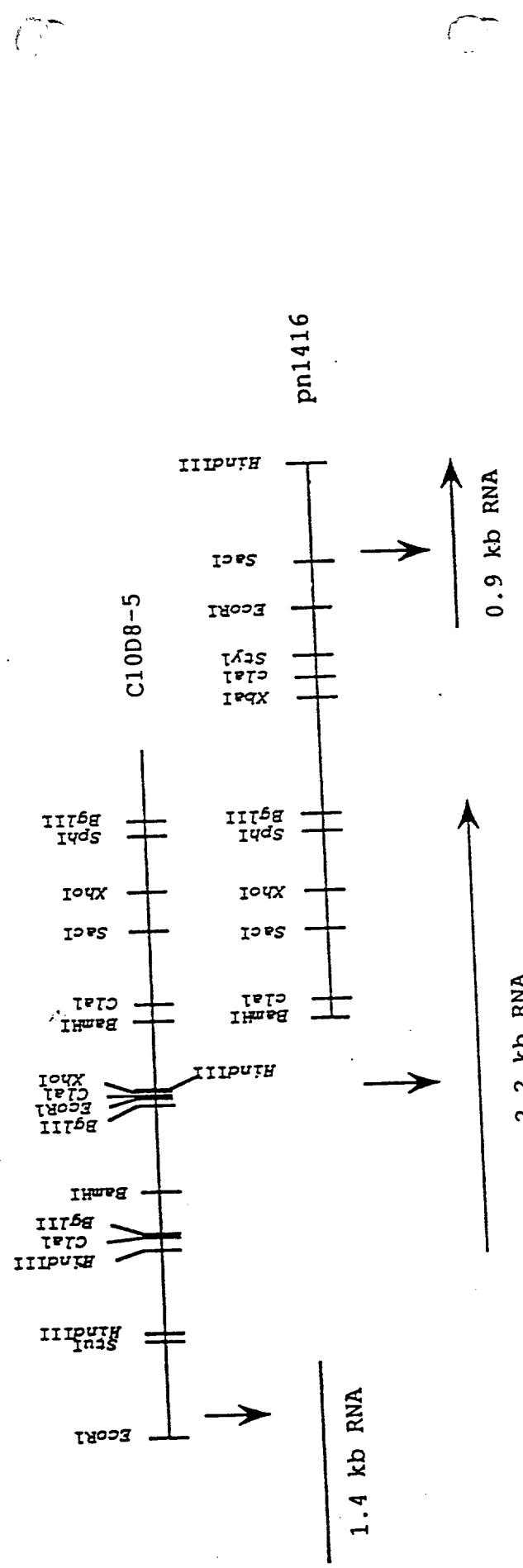


FIGURE 8

IV

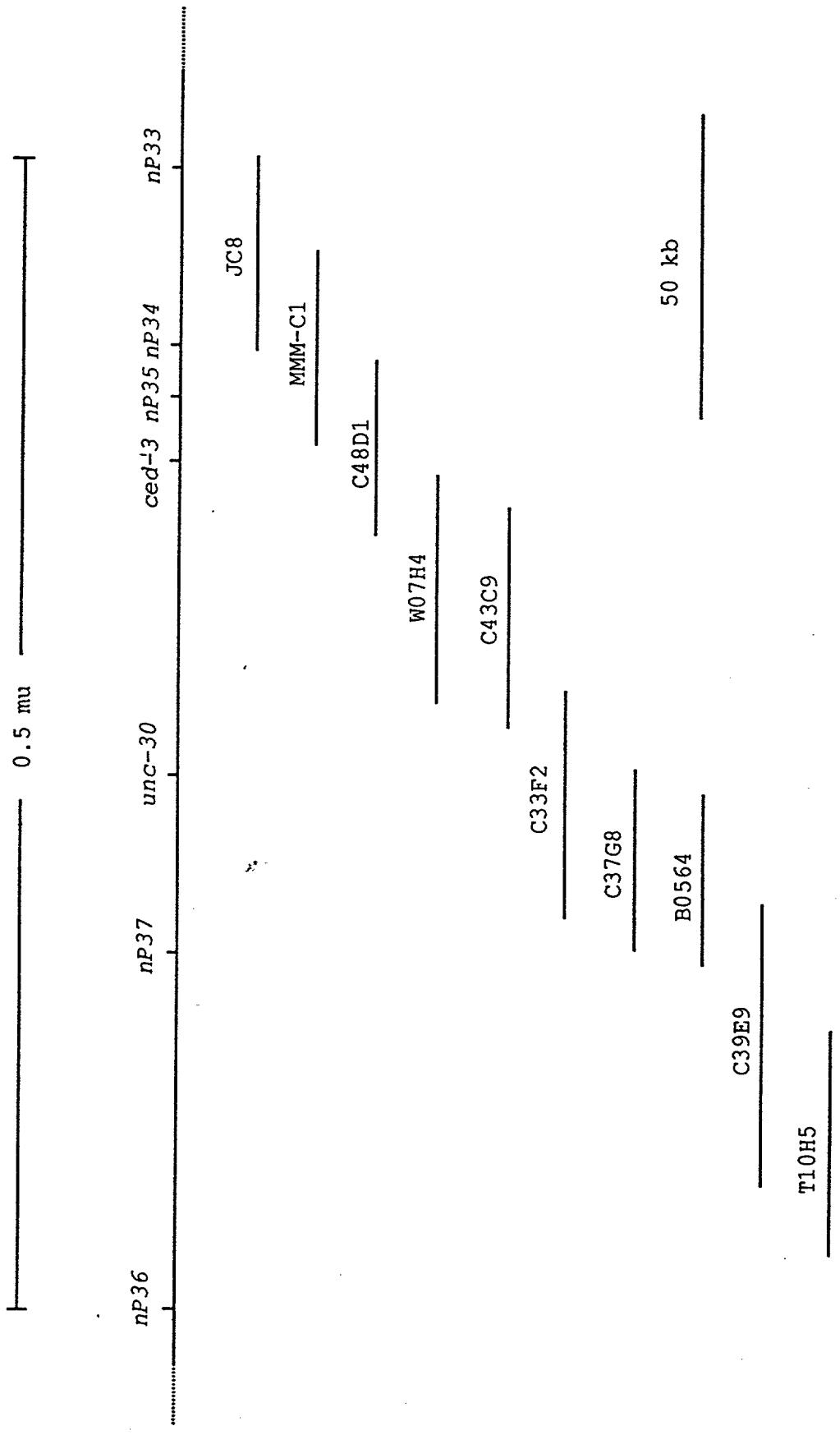


FIGURE 9

Figure 10 Summary of the experiments to localize *ced-3* gene within C48D1.